



Horizon Europe (HORIZON) - HORIZON-CL5-2024-D3-01-11

AI Testing and Experimentation Facility (TEF) for the energy sector – bringing technology to the market

Project no. 101172705



**Large-Scale Testing and Experimentation Facility (TEF) for Assessing,
Validating, and Enhancing AI-Powered Next-Generation Energy Solutions**

D2.1 - Landscape review analysis, EnergyGuard User requirement and specifications

By Hal Service

July 31, 2025

Workpackage: WP2 – Requirements, Specifications, Synergies and Alignment with European Ecosystem

Authors: HAL, ICCS, RDN, CIEMAT, CEA, CARTIF, REA, BER, MS, LXP, FORA, ENGREEN, SCCB

Status: Final

Date: 31/07/2025

Version: 1.00

Classification: Confidential

Disclaimer:

The EnergyGuard project is co-funded by the Horizon Europe Programme of the European Union under grant agreement No. 101172705. This document reflects only authors' views. The EC is not liable for any use that may be done of the information contained therein.

EnergyGuard Project Profile

Grant Agreement No.: 101172705

Acronym:	EnergyGuard
Title:	Large-Scale Testing and Experimentation Facility (TEF) for Assessing, Validating, and Enhancing AI-Powered Next-Generation Energy Solutions
URL:	http://www.energy-guard.eu
Start Date:	01/01/2025
Duration:	36 months

Partners

	EREVNITIKO PANEPISTIMIAKO INSTITOUTO SYSTIMATON EPIKOINONION KAI YPOLOGISTON (ICCS), Decision Support Systems Laboratory, DSSLab <u>Co-ordinator</u>	Greece
	CENTRO DE INVESTIGACAO EM ENERGIA REN - STATE GRID SA (RDN)	Portugal
	CENTRO DE INVESTIGACIONES ENERGETICAS MEDIOAMBIENTALES Y TECNOLOGICAS (CIEMAT)	Spain
	COMMISSARIAT A L ENERGIE ATOMIQUE ET AUX ENERGIES ALTERNATIVES (CEA)	France
	FUNDACION CARTIF (CARTIF)	Spain
	RIGA MUNICIPAL AGENCY "RIGA ENERGY AGENCY" (REA)	Latvia
	BLUENERGY REVOLUTION SCRL (BER)	Italy
	METASYMBOSE SAS (MS)	France
	LUXPROVIDE SA (LXP)	Luxembourg
	FUNDINGBOX ACCELERATOR SP ZOO (FBA)	Spain
	FORA FOREST TECHNOLOGIES SLL (FORA)	Spain
	FYLLA ANASTASIA (AF)	Greece
	ENGREEN SRL (ENGREEN)	Italy
	BUVNICIBAS VALSTS KONTROLES BIROJS (SCCB)	Latvia
	HAL SERVICE SPA (HAL)	Italy

Document History

Version	Date	Author (Partner)	Remarks
0.01	15/04/2025	Lorenzo Vandoni (HAL)	Table of Contents
0.03	04/05/2025	Panagiota Rempi, Theodosios Pountridis (ICCS)	ML and XAI in state-of-the-art analysis
0.04	11/05/2025	Anastasia Fylla (AF)	Regulation and Compliance in state-of-the-art analysis
0.05	16/05/2025	Marta Lafuente, Oscar Izquierdo (CIEMAT)	CIEMAT contributions in state-of-the-art analysis
0.06	22/05/2025	Nuno Pinho Da Silva, Inoussa Laouali (RDN)	RDN contributions in state-of-the-art analysis
0.07	28/05/2025	Frederik Pinel (LXP)	LXP contributions in state-of-the-art analysis
0.08	03/06/2025	Felix Bosio (CEA)	CEA contributions in state-of-the-art analysis
0.09	06/06/2025	Tommaso Ferrucci (ENGREEN)	ENGREEN contributions in state-of-the-art analysis
0.10	11/06/2025	Victor Serna, José Luis Harnandéz (CARTIF)	CARTIF contributions in state-of-the-art analysis
0.11	18/06/2025	Natalja Belska (REA)	REA contributions in state-of-the-art analysis
0.15	26/6/2025	Lorenzo Vandoni, Gabriele Balzano (HAL)	First Draft Handed for Internal Review
0.20	30/6/2025	Spiros Mouzakis, Sotiris Pelekis (ICCS)	Internal Review 1: Review report handed. Tracked changes and review comments integrated in document. The document needs serious restructuring.
0.30	1/7/2025	Spiros Mouzakis, Sotiris Pelekis (ICCS)	Substantial ToC and content restructuring
0.40	11/7/2025	Lorenzo Vandoni, Gabriele Balzano (HAL)	Second Draft Handed for Internal Review
0.50	18/7/2025	Lorenzo Vandoni, Gabriele Balzano (HAL)	Third Draft Handed for Internal Review
0.60	25/07/2025	Lorenzo Vandoni, Gabriele Balzano (HAL)	Fourth Draft Handed for Internal Review
0.70	28/07/2025	Frederico Lourenço (RDN)	Internal Review 2
1.00	31/07/2025	Spiros Mouzakis (ICCS)	Quality Control and Submission

Executive Summary

The EnergyGuard project, co-funded by the Horizon Europe Programme under Grant Agreement No. 101172705, aims to establish a large-scale Testing and Experimentation Facility (TEF) for the assessment, validation, and improvement of AI-powered energy solutions. The overarching goal of the project is to bridge the gap between innovative AI technologies and their real-world deployment in the energy sector, ensuring that these technologies are not only effective but also trustworthy, secure, and aligned with EU regulations and societal values.

This deliverable, D2.1, serves as a foundational document in the project, with two primary objectives: (1) to provide a comprehensive state-of-the-art review of AI technologies relevant to the energy sector, and (2) to collect and analyse detailed user requirements from a broad set of stakeholders. These insights will inform the design and implementation of the EnergyGuard TEF and its associated tools, services, and regulatory alignment strategies.

The document is structured into three main sections. After an introduction in Section 1, the second section presents a detailed state-of-the-art review, covering advanced AI methods—including machine learning, deep learning, large language models, and explainable AI—and their applications in energy systems such as smart grid operation, digital twins, predictive maintenance, and energy optimization. It also addresses critical enablers such as high-performance computing, big data management, cybersecurity, and ethical and regulatory frameworks at both EU and international levels.

The third section contains the requirements analysis, identifying common needs and expectations across stakeholder groups. This includes the results of tailored questionnaires directed at eight categories of stakeholders: AI developers, energy system stakeholders, legal experts, industry associations, investors, researchers, policy makers, and the general public. A ninth group—pilot use case leads—was included to gather technical specifications for real-world demonstration scenarios.

These requirements are then used to identify the Energyguard TEF specifications, providing a list of user personas, user stories and requirements.

Finally, this document contains two annexes, providing a short bibliography and the list of questionnaires that have been sent to the different types of stakeholders.

By combining technical, regulatory, and user-centered perspectives, this deliverable ensures that the EnergyGuard TEF will be grounded in current innovation while remaining aligned with the diverse needs of its future users and the broader European energy and digital policy landscape.

Table of Contents

1	Introduction	12
2	State of the art review	14
2.1	AI methods, services and tools.....	14
2.1.1	Machine Learning and Deep Learning in the Energy Sector.....	14
2.1.1.1	<i>Smart Grid Operation</i>	<i>15</i>
2.1.1.2	<i>Digital Twins for Buildings and Energy Efficiency</i>	<i>16</i>
2.1.1.3	<i>LLMs.....</i>	<i>18</i>
2.1.1.4	<i>AI Agents and Automation Workflows.....</i>	<i>20</i>
2.1.2	Explainable AI (XAI)	21
2.1.3	Federated Learning and Privacy-Preserving AI.....	25
2.2	AI applications for the energy domain	26
2.2.1	Digital twins and advanced simulations.....	26
2.2.1.1	<i>Architecture and classification of DTs.....</i>	<i>28</i>
2.2.1.2	<i>Digital Twin (DT) Applications in the Power Systems.....</i>	<i>29</i>
2.2.1.3	<i>Electrolysis system's Digital Twins</i>	<i>30</i>
2.2.2	AI for Smart Grid Management	32
2.2.2.1	<i>Machine Learning Approaches in Smart Grids</i>	<i>33</i>
2.2.2.2	<i>Deep Learning Advances in Smart Grid Management.....</i>	<i>34</i>
2.2.3	Optimization of Energy Demand and Resources.....	34
2.2.4	Predictive Maintenance and AI-Based Diagnostics.....	37
2.2.4.1	<i>Health management.....</i>	<i>38</i>
2.3	HPC and Big Data for AI-Driven Energy Solutions.....	41
2.3.1	High-Performance Computing (HPC).....	41
2.3.1.1	<i>High-Performance Computing in the Energy Sector.....</i>	<i>41</i>
2.3.1.2	<i>HPC for AI Overview</i>	<i>43</i>
2.3.1.3	<i>Infrastructure aspects</i>	<i>43</i>
2.3.1.4	<i>Software aspects</i>	<i>44</i>
2.3.1.5	<i>Access interface.....</i>	<i>44</i>
2.3.2	AI-HPC Convergence.....	44
2.3.3	Trusted and sovereign data sharing	45
2.3.4	Big Data management	48
2.4	Cybersecurity and trustworthy AI Assessment	49
2.4.1	Adversarial attacks and AI defence mechanisms.....	49
2.4.2	Bias, Fairness, and Explainability in AI for Energy	50
2.4.2.1	<i>Bias in Energy AI Systems</i>	<i>53</i>

2.4.2.2	<i>Fairness Considerations in Energy Systems</i>	54
2.4.2.3	<i>Explainability in AI for Energy</i>	56
2.4.3	Security and Governance Frameworks for AI	58
2.5	AI Regulation and compliance	60
2.5.1	International Regulations and compliance frameworks	60
2.5.1.1	<i>Background</i>	60
2.5.1.2	<i>Specific frameworks, regulations and guidelines related to AI</i>	62
2.5.1.3	<i>Charter of Fundamental Rights of the European Union</i>	62
2.5.1.4	<i>The Council of Europe Framework Convention on Artificial Intelligence</i>	63
2.5.1.5	<i>Artificial Intelligence Act</i>	64
2.5.1.6	<i>General Data Protection Regulation</i>	67
2.5.1.7	<i>Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union</i>	69
2.5.1.8	<i>Data Act</i>	69
2.5.1.9	<i>Directive on privacy and electronic communications/ e-privacy</i>	70
2.5.1.10	<i>NIS 2 Directive</i>	71
2.5.1.11	<i>Cybersecurity Act</i>	72
2.5.1.12	<i>Cyber Resilience Act (CRA)</i>	72
2.5.1.13	<i>Digital Single Market Directive</i>	73
2.5.1.14	<i>Information society Directive</i>	74
2.5.1.15	<i>Directive 96/9 on the legal protection of databases</i>	74
2.5.1.16	<i>Digital Services Act</i>	75
2.5.1.17	<i>European Declaration on Digital Rights and Principles for the Digital Decade</i>	75
2.5.1.18	<i>Approval of the content of the draft Communication from the Commission - Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act), European Commission, 6-2-2025</i>	76
2.5.1.19	<i>Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, EDPB</i> ...	78
2.5.1.20	<i>Methodology for the risk and impact assessment of artificial intelligence systems from the point of view of human rights, democracy and the rule of law (HUDERIA methodology), Council of Europe, Committee on Artificial Intelligence (CAI), 28-11-2024</i> ...	79
2.5.1.21	<i>Guidelines on prohibited artificial practices established by Regulation (EU) 2024/1689 (AI Act)</i>	80
2.5.2	Regulatory sandboxes and experimental frameworks	80

2.5.2.1	<i>Background</i>	80
2.5.2.2	<i>Regulations and other frameworks referring to regulatory sandboxes</i>	83
2.5.2.3	<i>Testing of high risks AI systems in real world conditions outside AI regulatory sandboxes</i>	86
2.5.2.4	<i>Union AI testing support structures</i>	87
2.5.2.5	<i>Difficulties of regulatory sandboxes in practice</i>	87
2.5.3	Ethical Considerations and Societal Impact of AI	88
2.5.3.1	<i>Ethics Guidelines for Trustworthy AI, High-Level Expert Group on Artificial Intelligence</i>	93
2.5.3.2	<i>Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment</i>	95
2.5.3.3	<i>AI Act</i>	97
2.5.3.4	<i>Recommendation on the Ethics of Artificial Intelligence, UNESCO</i>	97
2.5.3.5	<i>Recommendation of the OECD on Artificial Intelligence</i>	98
2.5.3.6	<i>Ethics by Design and Ethics of Use Approaches for Artificial Intelligence</i>	98
2.5.3.7	<i>Guidance regarding Ethical AI use in the energy sector, OFGEM</i> ..	99
2.5.4	Relevance to EnergyGuard	99

3 Requirements analysis 105

3.1	Questionnaires input	105
3.1.1	AI Developers	105
3.1.2	Energy Production and Energy Systems (EPES) value chain stakeholders..	112
3.1.3	Researchers and Academia.....	118
3.1.4	Other stakeholders	121
3.2	Pilot use cases	121
3.2.1	Pilot 1 – RDN	121
3.2.1.1	<i>Use Case 1 – Digital Twin of the Portuguese Transmission System Operation System</i>	121
3.2.1.2	<i>Use Case 2 – Automated Fault Analysis of Power System Events</i>	123
3.2.1.3	<i>Use Case 3 – Model Free Voltage Estimation</i>	125
3.2.2	Pilot 2 – CIEMAT – Microgrid-based Predictive Maintenance for Renewable Energy Equipment	127
3.2.3	Pilot 3 – CEA, CARTIF, BER	129
3.2.3.1	<i>Use Case 1 – CEA – MURPHY Digital Twin</i>	130
3.2.3.2	<i>Use Case 2 – CARTIF – AI-Powered Digital Twin for Hybrid Energy System Optimization</i>	132
3.2.3.3	<i>Use Case 3 – BER – DT-based MOSE Electrolyser Optimization</i> .	134

3.2.4 Pilot 4 – REA – Digital Twin Assisted, AI-driven Decision Support for Enhanced Building Energy Efficiency	136
3.2.5 Pilot 5 – ENGREEN – Enhancing performance of Renewable Energy Community Facilities through AI experiments.....	139
3.3 EnergyGuard TEF Specifications	144
3.3.1 User personas	145
3.3.2 User stories.....	147
3.3.3 Requirements.....	150
Annex I: Bibliography.....	157
Annex II: Requirements gathering - questionnaires	166
AI Developers	166
Energy Production and Energy Systems (EPES) value chain stakeholders.....	169
Lawyers and Legal system entities	171
Industry Associations & Technology Clusters	173
Investors	174
Researchers and Academia	175
Policy Makers & Standardisation Organisations (SDO).....	177
General Public.....	178
Pilot use cases	179

List of Tables

Table 1: Data Space Connectors Overview	48
Table 2: Expected relevance of the documents to EnergyGuard project	104
Table 3: User personas of the EnergyGuard TEF	146
Table 4: User personas of pilot applications.....	147
Table 5: User stories of the EnergyGuard TEF	148
Table 6: User stories of pilot applications	149
Table 7: Requirements for the EnergyGuard TEF	154
Table 8: Requirements for pilot applications	156
Table 9: Pilot use cases questionnaire.....	181

List of Acronyms

ADM	Automated Decision-Making
ADRA	AI Data Robotics Association
AEMEL	Anion-Exchange Membrane electrolysis
AI	Artificial Intelligence
ALE	Accumulated Local Effect

ALTAI	Assessment List for Trustworthy Artificial Intelligence
ANN	Artificial Neural Network
API	Application Programming Interface
ARIMA	AutoRegressive Integrated Moving Average
BAS	Building Automation Systems
BCBS	Basel Committee on Banking Supervision
BIM	Building Information Modelling
BoP	Balance of Plant
CAI	Committee on Artificial Intelligence
CIM	Common Information Model
CNN	Convolutional Neural Network
COBRA	Context-Based Analysis
COMTRADE	Common Format for Transient Data Exchange
CPU	Central Processing Unit
DER	Distributed Energy Resources
DL	Deep Learning
DMS	Distribution Management System
DRA	Distributed Resistance Analogy
DSA	Digital Services Act
DSMD	Digital Single Market Directive
DSO	Distribution System Operator
DT	Digital Twin
ECHR	European Convention on Human Rights
EDC	Eclipse Dataspace Connector
EDPB	European Data Protection Board
EIS	Electrochemical Impedance Spectroscopy
EMS	Energy Management Systems
ENISA	European Union Agency for Cybersecurity
EPC	Energy Performance Certificates
EPES	Energy Production and Energy Systems
ESCO	Energy Service Company
ESG	Environmental, Social, Governance
ETL	Extraction, Transformation and Loading
EU	European Union
FL	Federated Learning
FNN	Feedforward Neural Networks
GDPR	General Data Protection Regulation

GIS	Geographic Information System
GPAI	General-Purpose AI
GPR	Gaussian Process Regression
GPT	Generative Pre-trained Transformer
GPU	Graphics Processing Unit
GRU	Gated Recurrent Units
HLEG	High-Level Expert Group on Artificial Intelligence
HPC	High-Performance Computing
HVAC	Heating, Ventilation and Air Conditioning
ICE	Individual Conditional Expectation
ICT	Information and Communication Technology
IEC	International Electrotechnical Commission
IED	Intelligent Electronic Device
IoT	Internet of Things
ISD	Information Society Directive
KPI	Key Performance Indicator
LIME	Local Interpretable Model-Agnostic Explanations
LLM	Large Language Model
LSTM	Long Short-Term Memory
LV	Low Voltage
MAUP	Modifiable Areal Unit Problem
ML	Machine Learning
MLOps	Machine Learning Operations
MLP	Multi-Layer Perceptron
MOSE	MOdular Smart Electrolyzer
MPC	Model Predictive Control
MQTT	Message Queuing Telemetry Transport
MW	Mega Watt
NIS	Network and Information Security
NIST	National Institute of Standards and Technology
NGO	Public Sector Organization
NGSI-LD	Next Generation Service Interface – Linked Data
OFGEM	Office of Gas and Electricity Markets
OECD	Organisation for Economic Co-operation and Development
OS	Operating System
PdM	Predictive Maintenance
PDP	Partial Dependence Plots

PEM	Proton Exchange Membrane
PEMEL	Proton Exchange Membrane electrolysis
PHM	Prognostics and health management
PV	Photovoltaic
QPU	Quantum Processing Units
REC	Renewable Energy Communities
RECS	Residential Energy Consumption Survey
RES	Renewable Energy Sources
REST	Representational State Transfer
RAG	Retrieval-Augmented Generation
RNN	Recurrent Neural Network
RL	Reinforcement Learning
RIA	Risk and Impacts Assessment
SCADA	Supervisory Control And Data Acquisition
SED	Self-Encrypting Drives
SEP	Stakeholder Engagement Process
SHAP	Shapley Additive exPlanations
SME	Small and Medium Enterprises
SOC	Solid Oxide Cell
SOE	Solid Oxide Electrolysis
SOEC	Solid Oxyde Cell Electrolysis
SOFC	Solid Oxide Fuel Cell
SQL	Structured Query Language
SVM	Support Vector Machine
TCN	Temporal Convolutional Network
TEF	Testing and Experimentation Facility
TRISM	Trust, Risk, and Security Management
TSG	TNO Security Gateway
TSO	Transmission System Operator
UNESCO	United Nations Educational, Scientific and Cultural Organization
URL	Uniform Resource Locator
VPP	Virtual Power Plant
WIT	What-If Tool
WP	Work Package
XAI	Explainable Artificial Intelligence

1 Introduction

The **EnergyGuard** project, co-funded by the Horizon Europe Programme, aims to establish a **Large-Scale Testing and Experimentation Facility (TEF)** for assessing, validating, and improving Artificial Intelligence (AI)-powered next-generation energy solutions. The overarching ambition is to bridge the gap between advanced AI technologies and their deployment in the energy sector, ensuring that these technologies are trustworthy, secure, effective, and aligned with European values and regulatory frameworks.

This deliverable, **D2.1**, is a foundational document produced under **Work Package 2 (WP2) – Requirements, Specifications, Synergies and Alignment with European Ecosystem**. WP2 is responsible for laying the groundwork for the design of the EnergyGuard TEF by identifying state-of-the-art technologies, mapping regulatory and ethical considerations, and understanding user needs across various stakeholder groups.

Specifically, this deliverable covers three major tasks:

- A **landscape review** of state-of-the-art AI methods, tools, applications, and enabling technologies relevant to the energy domain;
- An **analysis of regulations and ethical frameworks**, including EU and international standards, guidelines, and compliance mechanisms related to AI in energy;
- A **comprehensive gathering and synthesis of user requirements**, collected via targeted questionnaires and interviews across multiple stakeholder categories, including TEF pilot leaders.

The results presented in this document are central to the development of the EnergyGuard TEF and are tightly integrated with other Work Packages. Insights from this deliverable feed into WP3 (Architecture and Reference Implementation), guide the TEF design activities in WP4, and support compliance and trustworthiness objectives in WP5 and WP6, particularly in relation to cybersecurity, data governance, and trustworthy AI.

The document is structured as follows:

- **Section 2** provides an **in-depth state-of-the-art review** of AI methods, services, and tools, as well as their applications in energy systems. It includes analysis of enablers such as High-Performance Computing (HPC), Big Data, cybersecurity, and relevant EU regulatory frameworks.
- **Section 3** presents the **user requirements analysis**, based on feedback from stakeholders and real-world pilot use cases. This includes user personas, user stories, and detailed TEF functional and technical specifications.
- The **Annexes** include the bibliography and the questionnaires used for requirements gathering.

By combining technical, regulatory, and user-centric perspectives, this deliverable ensures that the EnergyGuard TEF will be built on cutting-edge innovation while remaining responsive to the diverse needs of its future users and the broader European digital and energy ecosystem.

2 State of the art review

This section presents a comprehensive overview of the current state of the art in AI technologies relevant to the energy sector. It covers key methods, tools, and applications, along with enabling technologies such as HPC, big data, cybersecurity, and the regulatory and ethical frameworks shaping their deployment.

2.1 AI methods, services and tools

2.1.1 Machine Learning and Deep Learning in the Energy Sector

In recent years, the fields of Machine Learning (ML) and Deep Learning (DL) have experienced unprecedented growth, driving transformative changes across industries and reshaping how information is processed, understood, and acted upon.

At the forefront of this revolution are Large Language Models (LLMs), AI agents, and automation workflows—key components that are redefining the boundaries of artificial intelligence.

LLMs, such as GPT-based architectures, have demonstrated remarkable capabilities in natural language understanding and generation, enabling more human-like interaction between machines and users. Meanwhile, AI agents built on these models are evolving into autonomous systems capable of performing complex tasks, making decisions, and collaborating in dynamic environments. Automation workflows powered by these technologies are streamlining processes, increasing efficiency, and opening new possibilities in domains ranging from software development to scientific research.

One such domain significantly impacted by these advancements is the energy sector. Here, various applications often involve the measurement and collection of several types of data from smart meters and other sensors. This data is used to make predictions about system states or the future behaviour of energy consumers and producers. Many applications such as state estimation can benefit¹. However, these equations can be very time consuming to solve, or they may not cover a particular application. In these cases, purely or partially data-driven methods can be utilised.

For years, traditional (statistical) models were the first choice of data-driven methods, especially for energy consumption / production forecasting. Models like ARIMA² and exponential smoothing³ were used. However, the advent of ML and DL methods has challenged this dominance. Examples of these models include LSTMs⁴, Transformers⁵, Decision trees⁶ and MLPs⁷. These kinds of models can handle

¹ <https://pdfs.semanticscholar.org/f5c7/a4e353c154eccd8909cd26dbbedca0c2d767.pdf>

² <https://ieeexplore.ieee.org/document/5401272>

³ <https://ieeexplore.ieee.org/document/4074410>

⁴ <https://doi.org/10.1162/neco.1997.9.8.1735>

⁵ https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

⁶ <https://doi.org/10.1201/9781315139470>

⁷ <https://www.nature.com/articles/323533a0>

the increasing complexity and size of the data better⁸. This deliverable will further detail the use of ML and DL in a wide range of energy applications.

This chapter explores the current state of the art in ML and DL, with a focus on LLMs, AI agents, intelligent automation, and their transformative impact on energy applications.

2.1.1.1 *Smart Grid Operation*

ML and DL models have several applications in smart grid operation, to forecast energy demand and generation in real time. The smart grid can then dynamically balance supply and optimize the utilization of its generating capacity. Various methods have been utilised to forecast energy demand and production in smart grids, ranging from MLPs⁹ and LSTMs¹⁰, to LightGBM¹¹ and other Decision Trees¹². Other more promising recent methods include the use of Transformers¹³ and global forecasting models in general¹⁴ that are trained using all the time series of the dataset instead of making a single model for each one. This allows them to take advantage of the relationships between all consumers and producers. Considering these advantages, EnergyGuard aims

EnergyGuard ML models can be employed to anticipate and localize emerging faults and speed up the recovery so the grid can ride through disturbances and swiftly restore reliable service. Recent studies show increasing use of both ML and DL methods to predict outages and for fault detection. For example, multi-agent deep reinforcement learning from Ruiqi et. al.¹⁵ autonomously restores the affected loads after extreme events, using a dynamic-agent network to cope with a changing topology. An XGBoost based method by Ghasemkhani et al.¹⁶ predicts outage durations with 98 % accuracy on utility data, allowing operators to manage outages and communicate with customers better. Guohao¹⁷ trains an MLP model that analyses frequency-response-analysis curves and achieves >99 % accuracy identifying winding-fault degree in power-grid transformers. Tianqiao et. al.¹⁸ builds a stacked bi-directional LSTM followed by an MLP layer, that predicts storm-driven outages up to 120 h ahead, outperforming

⁸ <http://www.deeplearningbook.org>

⁹ <https://www.mdpi.com/1996-1073/15/4/1295>, <https://doi.org/10.1016/j.segan.2023.101171>

¹⁰ 10.21203/rs.3.rs-5458984/v1, <https://www.frontiersin.org/journals/energy-research/articles/10.3389/fenrg.2023.1116465/full>

¹¹ https://www.researchgate.net/publication/380832440_A_transfer_learning-based_hybrid_model_with_LightGBM_for_smart_grid_short-term_energy_load_prediction, 10.1109/ACCESS.2022.3192011

¹² <https://doi.org/10.3390/en15207547>, <https://doi.org/10.3390/en17081926>

¹³ <https://arxiv.org/abs/2310.17477>, <https://doi.org/10.1016/j.tej.2024.107370>

¹⁴ <https://doi.org/10.1109/TSG.2023.3264525>, <https://www.sciencedirect.com/science/article/pii/S0306261921003986>

¹⁵ <https://doi.org/10.1016/j.apenergy.2024.123625>

¹⁶ <https://doi.org/10.3390/s24134313>

¹⁷ <https://arxiv.org/pdf/2406.19623>

¹⁸ <https://ieeexplore.ieee.org/document/10328277>

shallower RNN baselines on U.S. utility data. Finally, Razieh et. al.¹⁹ propose The Informer—a transformer-based model—that anticipates outage scenarios and offers very small response time that can be utilised during an emergency. In EnergyGuard, for Pilot 1 (RDN), EnergyGuard plans leverage state-of-the-art ML and DL methods along with a signal processing technique to detect and classify faults in power systems by analysing the COMTRADE files generated by Intelligent Electronic Devices (IED). This approach will improve grid resilience as automatic and timely fault detection can lead to faster recovery and localisation of the fault.

2.1.1.2 *Digital Twins for Buildings and Energy Efficiency*

In the field of **digital twins for building energy efficiency**, AI and ML are increasingly being leveraged to enhance energy management and decision-making. Popular techniques and models are used for the modelling step of a digital twin, with supervised machine learning algorithms being frequently employed. These include regression approaches (e.g., linear, polynomial regression), classification approaches (e.g., support vector machines, decision trees), and probabilistic algorithms (e.g., Artificial Neural Networks (ANN), Naive Bayes). Artificial Neural Networks (ANN) were identified as the most frequently applied in the reviewed literature²⁰. Other researchers also utilised a hybrid approach, combining physics-based (white-box) and statistical (blackbox) models²¹. Furthermore, support vector machines and decision trees are also used^{22 23}, along with Bayesian networks^{24 25}. In EnergyGuard, for Pilot 4 (REA), EnergyGuard aims to use AI and machine learning models to energy efficiency within the Riga multi-apartment residential buildings. By integrating data-driven approaches, EnergyGuard will enhance energy management and decision-making in this specific urban setting.

Electrolyser Prognostics and Integration

Regarding electrolyser prognostics and intelligent maintenance, the literature highlights the use of various machine learning (ML) and deep learning (DL) techniques to enhance fault detection, performance prediction, and durability. Supervised learning is commonly employed, with artificial neural networks (ANNs), including long short-term memory (LSTM) networks, used for fault detection and localisation in power electronics, modelling the aging and degradation of PEM electrolytic stacks, and predicting multiple design parameters of electrolytic cells²⁶). Other supervised techniques such as

¹⁹ <https://doi.org/10.1016/j.ijepes.2025.110627>

²⁰ <https://doi.org/10.3390/en14082338>

²¹ <https://doi.org/10.1145/3447555.3466585>

²² <https://doi.org/10.1016/j.enbuild.2022.111988>

²³ <https://doi.org/10.1109/RUSAUTOCON.2019.8867611>

²⁴ <https://doi.org/10.3390/en14082338> , <https://doi.org/10.3390/environsciproc2021012010>

²⁵ <https://www.mdpi.com/1996-1073/15/19/7002>

²⁶ <https://doi.org/10.1109/ECCE47101.2021.9595704>

<https://doi.org/10.1002/cite.202000092>

<https://doi.org/10.1016/j.ijhydene.2022.01.158>

<https://doi.org/10.3390/en15186657>

polynomial and logistic regression are also applied for design parameter prediction (e.g., study²⁷), while support vector machines (SVM) and Gaussian process regression (GPR) have been used in prognostics and health management (PHM) to predict load voltage as a state-of-health indicator²⁸. Hybrid AI approaches, combining fuzzy logic with neural networks, have further been explored to improve electrolyser durability²⁹. In the area of optimisation, reinforcement learning (RL) techniques, including feature selection algorithms, have been utilized to identify optimal predictive features for maintenance and operation³⁰. For electrolyser prognostics and intelligent maintenance, EnergyGuard will apply a combination of machine learning and deep learning techniques, including ANNs and LSTM networks, to enhance fault detection, performance prediction, and durability. These models will be tested in real-world scenarios to optimize maintenance strategies and improve the operational lifespan of electrolyser systems.

Renewable Energy Communities (REC)

The primary goal of AI in Renewable Energy Communities (RECs) is to enhance system reliability, flexibility, and efficiency, supporting the transition toward carbon neutrality. Supervised learning techniques, particularly support vector regression and hybrid models, have been widely applied for forecasting heating, cooling, and electrical loads in buildings³¹. Deep learning methods, including deep belief networks and recurrent neural networks, are commonly used for similar demand prediction tasks, especially in electricity and cooling load forecasting³². Reinforcement learning, with approaches such as actor-critic algorithms, plays a significant role in real-time energy management and online optimal scheduling for smart buildings³³. Unsupervised learning is primarily applied for clustering applications within energy systems, while machine learning has also proven useful in predicting the performance of renewable energy technologies, such as photovoltaic power output, through surrogate modelling³⁴. In EnergyGuard, AI will be applied within RECs to optimize system reliability and efficiency. Techniques such as support vector regression, deep learning methods like recurrent neural networks, and

²⁷ <https://doi.org/10.3390/en15186657>

²⁸ <https://doi.org/10.1016/j.egyai.2023.100251>

²⁹ <https://doi.org/10.1016/j.ijhydene.2022.01.158>

³⁰ <https://doi.org/10.1016/j.camwa.2013.07.019>

<https://doi.org/10.3390/s23010098>

<https://doi.org/10.3390/a16120541>

³¹ <https://doi.org/10.1016/j.energy.2020.118676>

<https://doi.org/10.1016/j.rser.2017.05.249>

<https://doi.org/10.1016/j.apenergy.2017.03.070>

³² <https://doi.org/10.1016/j.energy.2018.01.180>

<https://doi.org/10.1016/j.apenergy.2017.12.051>

³³ <https://doi.org/10.1109/TSG.2018.2834219>

<https://doi.org/10.1016/j.epsr.2020.106543>

³⁴ <https://doi.org/10.1016/j.renene.2019.11.037>

<https://doi.org/10.1016/j.egyai.2022.100187>

reinforcement learning can be used for load forecasting and real-time energy management, ensuring efficient operation of smart buildings and renewable energy systems.

2.1.1.3 LLMs

Large Language Models (LLMs) are a class of deep learning models trained on vast *corpora* of textual data, enabling them to perform a wide range of natural language processing (NLP) tasks. These models, based primarily on the Transformer architecture, have achieved remarkable performance in areas such as language generation, translation, summarization, and question answering.

The Transformer architecture introduced by Vaswani et al. in 2017 catalysed a significant evolution in NLP. Subsequent models such as BERT, GPT, and T5 adopted and extended this architecture, emphasizing scalability and parallelism. GPT models, characterized by their autoregressive structure, have enabled highly fluent and contextually aware text generation³⁵. Comparative evaluations highlight GPT-4's superior performance in areas like creativity, translation accuracy, and handling imperfect prompts when compared to models such as BERT³⁶.

LLMs have become central to numerous applications, such as:

- Text Generation and Dialogue Systems: GPT-based models' power conversational AI and assistive writing tools, excelling in both fluency and coherence.
- Information Retrieval and Summarization: LLMs are used to condense long-form content while maintaining semantic relevance.
- Creative Applications: From poetry to marketing content, LLMs enhance human creativity through rapid ideation.
- Specialized Control Systems: LLMs are being integrated into robotics workflows, enabling natural language control of drones and autonomous agents.

Despite their success, LLMs face key challenges:

- Hallucinations: Tendency to generate plausible but factually incorrect statements.
- Bias and Fairness: Risk of encoding and amplifying harmful societal biases from training data.
- Interpretability: The internal mechanisms of models like BERT remain difficult to interpret, although some studies have explored layer-wise behaviour³⁷.
- Resource Intensity: Training and deploying LLMs demands immense computational resources, often limiting access to large institutions.

Recent research has emphasized:

- Prompt Engineering and Fine-Tuning: Optimizing inputs to elicit specific behaviours without retraining the model³⁸.

³⁵ <https://consensus.app/papers/the-architecture-of-language-understanding-the-mechanics-ferraris-audrito/37e3ae7d2b855c0fbded82f36d971438/>

³⁶ <https://consensus.app/papers/unmasking-large-language-models-by-means-of-openai-gpt4-and-zahid-joudar/265952bb45d85ff1bbb353bebda38c8f/>

³⁷ <https://consensus.app/papers/exploring-narrative-clustering-in-large-language-models-a-banerjee-schilling/377c811e893c59a2bb0f51ac27c5e066>

³⁸ <https://consensus.app/papers/unmasking-large-language-models-by-means-of-openai-gpt4-and-zahid-joudar/265952bb45d85ff1bbb353bebda38c8f/>

- Retrieval-Augmented Generation (RAG): Enhancing factual accuracy by coupling LLMs with external knowledge bases³⁹.
- Domain-Specific Evaluation Frameworks: Especially in critical fields like medicine, where precision and ethical alignment are vital⁴⁰.

According to Majumer S, et al.⁴¹, “LLMs offer promising value to the electric energy sector, thanks to their ability to interpret human prompts and alleviate sensory overload, especially providing near real-time guidance in managing extreme weather events and risks associated with diverse sources of uncertainty”.

Using the Generative Pre-trained Transformer (GPT) models, the authors identify a set of capabilities and limitations of applying LLMs in the Electric Energy Sector:

- Capabilities:
 - Correlation analysis for power systems.
 - Wildfire risks recognition.
 - Equipment damage detection.
 - On-site hazard recognition.
 - Document analysis for power systems.
 - Load forecast.
- Limitations:
 - Price forecast, due to the limited LLM adaptation to handle physics and potential exposure to cyber security and privacy threats.
 - Power flow related problems, due to challenges in domain-specific data availability and processing, limited adaptation to handling physics and potential exposure to cyber security and privacy threats.
 - Ensuring safe power system operation due to challenges in domain specific data availability and processing, as well as lack of safety guardrails. Potential security concerning privacy invasion through LLMs, deterioration of LLM performance, semantic divergence and denial of service, are reviewed in Ruan J., et al.⁴².

In Huang C., et al⁴³, the GPT-4 and GPT-4 Vision (GPT-4V) are exploited to address power system tasks. The authors design specific prompts and pipelines, thus using LLMs to:

- Solve simplified OPF by only interacting with prompts without using optimisation solvers

³⁹ <https://consensus.app/papers/the-architecture-of-language-understanding-the-mechanics-ferraris-audrito/37e3ae7d2b855c0fbded82f36d971438/>

⁴⁰ <https://consensus.app/papers/a-comprehensive-survey-on-evaluating-large-language-model-huang-tang/b3804f50f197532e823bc3c9e06c4e33/>

⁴¹ Majumer S, et al., “EXPLORING THE CAPABILITIES AND LIMITATIONS OF LARGE LANGUAGE MODELS IN THE ELECTRIC ENERGY SECTOR”, 2024 ([Exploring the capabilities and limitations of large language models in the electric energy sector: Joule](#))

⁴² Ruan J., et al., “Applying Large Language Models to Power Systems: Potential Security Threats”, 2024.

⁴³ Huang C., et al., “Large Foundation Models for Power Systems”, 2024 ([Large Foundation Models for Power Systems | IEEE Conference Publication | IEEE Xplore](#))

- Intermediate user interactions with EV charging control algorithms, so that optimal scheduling can be made based on users' preferences expressed through natural language
- Situation awareness by exploiting GPT-4V in the analysis of satellite images to detect wildfires, although still with lower accuracy than standard machine learning models
- Energy knowledge retrieval by using LLM to answer both generic and technical questions, where the use of retrieval augmented generation with the GPT-4 improved the domain specific answers to the technical questions.

On March 31, 2025, the European Commission hosted the webinar "GenerativeAI4EU; Accelerating Quick Wins for Smart Grids" in collaboration with the AI Data Robotics Association (ADRA). In this webinar a couple of LLM applications were highlighted⁴⁴:

- Enexis integrated LLM into workflows, such as summarising calls, generating documentation, or querying procedures
- The Helmholtz Association are exploiting LLM to analyse scientific literature, orchestrate lab workflows and identify novel combinations of materials.

2.1.1.4 AI Agents and Automation Workflows

The rapid evolution of AI has profoundly transformed the energy sector, enabling the development of intelligent agents and automation workflows that optimize operations, enhance decision-making, and improve sustainability. As detailed in recent research, AI-based modelling techniques—such as machine learning, deep learning, natural language processing, and expert systems—are at the forefront of creating autonomous systems tailored for energy applications.

AI agents are autonomous software entities capable of perceiving their environment, reasoning⁴⁵ and executing actions without human intervention. In energy applications, these agents facilitate real-time monitoring, predictive maintenance, and adaptive control of energy grids and renewable resources. For instance, AI-powered agents can analyse vast amounts of sensor data from smart meters, wind turbines, or solar panels to detect anomalies, forecast energy production, and optimize resource allocation. This autonomous decision-making reduces operational costs and minimizes downtime, contributing to more resilient and efficient energy infrastructures.

Automation workflows leverage AI techniques to streamline complex processes within the energy supply chain. These workflows encompass data collection, processing, modelling, and actuation, often integrated into a cohesive system. For example, in smart grids, automation workflows use AI algorithms to balance load demand, predict consumption patterns, and manage distributed energy resources dynamically. ML models can forecast peak usage periods, enabling utilities to pre-emptively adjust generation and storage strategies. Similarly, in renewable energy farms, AI-driven automation can optimize blade angles in wind turbines or track sun movement in solar panels to maximize energy capture.

AI models categorized as analytical, functional, and interactive AI play critical roles in energy workflows. Analytical AI extracts insights from sensor data to inform strategic decisions, such as capacity planning or price setting. Functional AI automates actions like switching transformers or controlling energy

⁴⁴ Post event report 2025 ([ADRAPostEvent GenerativeAI_May2025.pdf - Google Drive](#))

⁴⁵ González-Briones, Alfonso, et al. "Energy optimization using a case-based reasoning strategy." *Sensors* 18.3 (2018): 865.

dispatch based on real-time data. Interactive AI, including chatbots or virtual assistants, enhances customer engagement and service management by automating inquiries and providing personalized energy consumption advice.

DL architectures, such as Convolutional Neural Networks (CNNs) and recurrent neural networks (RNNs), enable advanced pattern recognition in energy data, supporting fault detection and predictive maintenance. Data mining and knowledge discovery techniques further extract valuable insights from historical data, aiding in trend analysis and scenario planning.

Despite significant advancements, implementing AI agents and automation workflows in energy remains challenging. Data quality and integration from heterogeneous sources are critical concerns. Developing scalable, interpretable, and robust AI models to handle the dynamic and uncertain nature of energy systems is essential. Moreover, ensuring cybersecurity and safeguarding privacy within automated energy networks are vital research areas.

In EnergyGuard, several AI workflows will be developed for all pilot applications using Dagster workflow orchestration engine. Example workflow applications include workflows for electrical load forecasting that covers all the tasks of the AI value chain, from data ingestion, preprocessing and harmonisation, model training, evaluation and deployment, task monitoring and error handling, scheduling and dynamic triggering. These workflow applications will be adapted to the needs of several pilots (e.g. Pilot 1 - RDN), utilising their data and satisfying their needs. Another example includes workflows for development, training and deployment of energy efficiency optimisation algorithms for Pilot 4 (REA). Furthermore, as the project moves forward, agentic AI solutions will be examined to enhance decision-making and user experience, on EnergyGuard's AI applications. Significant attention will be paid on self-improving AI models (via continual learning) using AI agents to evaluate and refine predictions over time providing feedback and alerts in near real-time.

2.1.2 Explainable AI (XAI)

Explainable Artificial Intelligence (XAI) is essential for enhancing transparency, accountability, and trust in AI systems, particularly as complex machine learning (ML) models become increasingly prevalent in high-stakes domains. As AI is integrated into critical sectors and models grow in complexity, the ability to understand and justify AI-driven predictions becomes a fundamental requirement for responsible use. XAI addresses this need by offering a set of processes and methods that render decision-making comprehensible to human users, thereby enabling informed oversight, regulatory compliance, and ethical implementation. It facilitates the detection of bias, supports debugging to improve system design, and ensures that model outcomes are grounded in meaningful, interpretable factors. Ultimately, XAI ensures that AI systems are not only accurate but also understandable, fair, and aligned with human values, key attributes for fostering responsible and trustworthy AI in real-world applications.

As XAI is evolving rapidly, various approaches have been developed to integrate explainability into different model types and application contexts. Certain ML models, such as decision trees and linear regression, are inherently interpretable due to their transparent structure and comprehensible decision-making processes. However, many ML models operate as black boxes, where increased complexity results in reduced explainability. In such cases, post-hoc explainability methods are employed after model training. These techniques are typically classified as either model-specific, which are tailored to particular types of models, or model-agnostic, which analyse the relationship between inputs and outputs without requiring access to the model's internal structure. Additionally, XAI techniques are

distinguished as local or global. Local approaches focus on explaining individual predictions, whereas global methods aim to capture the overall influence of features across the entire dataset on average.

A brief yet comprehensive overview of representative XAI tools and services follows, highlighting the current state-of-the-art landscape of the field. Specific applications of XAI within the energy domain are described in Section 1.4.2.

Shapley Additive exPlanations (SHAP)⁴⁶ is a well-established model-agnostic explainability method grounded in game theory's Shapley values. It quantifies the contribution of each feature to a model's prediction, offering both local and global explanations. SHAP assigns each feature an importance value that reflects its positive or negative impact on the model's output, helping identify the most influential features. Variants like TreeSHAP and KernelSHAP provide optimized implementations for different model types. SHAP is widely adopted for its solid theoretical foundation and strong visual interpretability.

Local Interpretable Model-Agnostic Explanations (LIME)⁴⁷ is a model-agnostic technique developed to provide local, instance-specific explanations for predictions generated by black-box ML models. The method operates by introducing slight perturbations to the input data and training a simpler interpretable model to approximate the behaviour of the complex initial one. This surrogate model provides insights into which features influenced the initial model's prediction. LIME supports various data modalities, including tabular, textual, and visual inputs, and produces explanations that are easily interpretable by human users.

Partial Dependence Plots (PDPs)⁴⁸ visualize the relationship between one or two features and the predicted outcome of a ML model. A PDP demonstrates how the model's prediction changes as a single feature varies, while holding other features constant. This approach is straightforward to implement and interpret, as it provides clear, intuitive visual representations that facilitate a deeper understanding of the influence of specific features on the model's predictions.

Accumulated Local Effects (ALEs)⁴⁹ show the average influence of features on model predictions offering a faster, unbiased alternative to PDPs.

⁴⁶ S. M. Lundberg, S. I. Lee, A unified approach to interpreting model predictions, in: Advances in Neural Information Processing Systems, Vol. 2017-December, 2017.

⁴⁷ M. T. Ribeiro, S. Singh, C. Guestrin, "Why Should I Trust You?": Explaining the Predictions of Any Classifier, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, Association for Computing Machinery, New York, NY, USA, 2016, p. 1135–1144. doi:10.1145/2939672.2939778.

⁴⁸ C. Molnar, Interpretable Machine Learning, 2nd Edition, 2022. <https://christophm.github.io/interpretable-ml-book>

⁴⁹ D. W. Apley, J. Zhu, Visualizing the effects of predictor variables in black box supervised learning models, Journal of the Royal Statistical Society Series B: Statistical Methodology 82 (4) (2020) 1059–1086, doi:10.1111/rssb.12377.

Individual Conditional Expectation (ICE)⁵⁰ Curves are the local equivalent to PDPs displaying one line per instance. ICE curves improve interpretability by highlighting heterogeneous relationships but are limited to one feature per plot and may be misleading if features are correlated.

Counterfactual Explanations⁵¹ are a model-agnostic technique in explainable AI that identify the minimal changes to input features required to alter a model's prediction. By describing how specific modifications can lead to a different outcome, they offer actionable insights and enhance understanding of model behaviour. This approach is especially effective for black-box models and binary classification tasks.

Deep Learning Important Features (DeepLIFT)⁵² is a model-specific interpretability method for deep neural networks that attributes the output prediction to input features by backpropagating contribution scores. It compares neuron activations to reference activations and efficiently computes both positive and negative contributions in a single backward pass, revealing feature dependencies often missed by other techniques.

Anchor Explanations⁵³ are a form of local explanations that provides simple if-then rules identifying a set of feature conditions under which a ML model consistently outputs the same prediction. These rules "anchor" the prediction locally, meaning that if the anchor conditions are satisfied, variations in other features do not significantly affect the outcome. Anchors are precise, interpretable, and particularly useful for generating transparent, human-understandable explanations.

InterpretML⁵⁴ is an open-source Python toolkit by Microsoft that provides a unified framework supporting both glassbox models - inherently interpretable models (e.g., linear models, rule lists) - and black-box XAI methods (e.g. LIME, PDP). The package offers a wide range of techniques with interactive visuals that allow users to easily evaluate model attributes such as performance, global and local features.

⁵⁰ A. Goldstein, A. Kapelner, J. Bleich, E. P. and, Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation, *Journal of Computational and Graphical Statistics* 24 (1) (2015) 44–65. <https://doi.org/10.1080/10618600.2014.907095>

⁵¹ M. T. Keane, B. Smyth, Good counterfactuals and where to find them: A case-based technique for generating counterfactuals for explainable AI (XAI), in: I. Watson, R. Weber (Eds.), *Case-Based Reasoning Research and Development*, Springer International Publishing, Cham, 2020, pp. 163–178.

⁵² A. Shrikumar, P. Greenside, A. Kundaje, Learning important features through propagating activation differences (2019) arXiv:1704.02685.

⁵³ M. T. Ribeiro, S. Singh, C. Guestrin, Anchors: High-precision model-agnostic explanations, *Proceedings of the AAAI Conference on Artificial Intelligence* 32 (1) (Apr. 2018). doi:10.1609/aaai.v32i1.11491.

⁵⁴ H. Nori, S. Jenkins, P. Koch, R. Caruana, InterpretML: A Unified Framework for Machine Learning Interpretability, arXiv preprint arXiv:1909.09223 (2019).

AI Explainability 360 (AIX360)⁵⁵ is an open-source Python toolkit developed by IBM to enhance the explainability of ML models. It includes a wide range of algorithms tailored for different model types and supports domain-specific applications. AIX360 also provides tools to promote fairness and transparency, making it suitable for high-stakes or regulated environments.

AI Fairness 360 (AIF360)⁵⁶ is a complementary open-source Python package designed to detect, evaluate, and mitigate bias in ML models. It includes a comprehensive set of fairness metrics and mitigation algorithms applicable to both data and models, supporting the development of more trustworthy AI systems.

What-If Tool (WIT)⁵⁷, developed by Google, is a user-friendly interface for explaining ML models. It enables interactive analysis of model predictions, feature importance, fairness metrics, and performance under hypothetical scenarios. Users can edit input data manually or programmatically and visualize the effects in real time, supporting comprehensive model understanding across diverse data subsets.

ELI5 (Explain Like I'm 5)⁵⁸ is a Python package used to debug machine learning classifiers and explain their predictions. It supports various types of models, such as tree-based models, linear models, and deep learning models, and works with multiple machine-learning frameworks, including Keras. ELI5 provides easy-to-understand, human-readable explanations, making machine learning more accessible to a wider audience.

Captum⁵⁹ is an open-source, extensible library built on PyTorch that provides a wide range of algorithms for model interpretability.

Alibi⁶⁰ is an open-source Python library for inspecting and interpreting machine learning models. It provides a diverse set of explanation algorithms, such as anchors, ALE, and counterfactuals, primarily focused on black-box models and local interpretability. Alibi supports various data types and includes methods suitable for both classification and regression tasks.

⁵⁵ V. Arya, R. K. E. Bellamy, P.-Y. Chen, A. Dhurandhar, M. Hind, S. C. Hoffman, S. Houde, Q. V. Liao, R. Luss, A. Mojsilovic, S. Mourad, P. Pedemonte, R. Raghavendra, J. Richards, P. Sattigeri, K. Shanmugam, M. Singh, K. R. Varshney, D. Wei, Y. Zhang, One Explanation Does Not Fit All: A Toolkit and Taxonomy of AI Explainability Techniques (2019).

⁵⁶ R. K. E. Bellamy, K. Dey, M. Hind, S. C. Hoffman, S. Houde, K. Kannan, P. Lohia, J. Martino, S. Mehta, A. Mojsilovic, S. Nagar, K. N. Ramamurthy, J. Richards, D. Saha, P. Sattigeri, M. Singh, K. R. Varshney, Y. Zhang, AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias (Oct. 2018), <https://arxiv.org/abs/1810.01943>

⁵⁷ What-if-tool <https://pair-code.github.io/what-if-tool/explore/>

⁵⁸ ELI5: A library for debugging/inspecting machine learning classifiers and explaining their predictions <https://eli5.readthedocs.io/>

⁵⁹ Captum Model Interpretability for PyTorch <https://captum.ai/>

⁶⁰ J. Klaise, A. V. Looveren, G. Vacanti, A. Coca, Alibi Explain: Algorithms for Explaining Machine Learning Models, Journal of Machine Learning Research 22 (181) (2021) 1–7

2.1.3 Federated Learning and Privacy-Preserving AI

As the energy sector becomes increasingly data-driven, the demand for collaborative AI models that respect data privacy and sovereignty has grown considerably. Federated Learning (FL) has emerged as a key technique for developing smart, privacy-friendly energy systems, especially in contexts like smart grids and microgrids. FL enables multiple stakeholders—such as energy producers, consumers, and utility operators—to collaboratively train machine learning models without directly sharing sensitive data. This decentralized approach aligns naturally with privacy-preserving goals, making it a critical component of modern AI systems that must comply with data protection laws like GDPR.

Federated Learning (FL) is a decentralized machine learning paradigm in which model training is performed collaboratively across multiple devices or servers holding local data samples, without exchanging the raw data itself. Instead of sending data to a central server, each client trains the model locally and only shares model updates (such as gradients or weights), which are then aggregated to form a global model. This model is particularly valuable in scenarios where data privacy, security, or bandwidth limitations make traditional centralized learning impractical.

Privacy-Preserving AI is a broader concept that encompasses various techniques—FL among them—designed to protect sensitive data throughout the AI lifecycle. This includes methods like differential privacy, homomorphic encryption, and secure multi-party computation, which are increasingly integrated into FL frameworks to enhance privacy. These technologies ensure that powerful AI systems can learn from and interact with data without exposing it or compromising confidentiality.

FL is widely adopted in sectors like healthcare, finance, mobile systems, industrial IoT, and energy, where direct data sharing is risky or legally restricted. Its decentralized nature—keeping data on local devices and only sharing model updates—makes it a leading approach for privacy-preserving AI, aligning with regulations such as GDPR.

Despite its advantages, FL also faces **several challenges**. One major issue is security (Ekellem, 2023⁶¹). Another challenge is data heterogeneity, where non-IID (non-independent and identically distributed) data across clients can degrade model performance. To address this, personalized FL techniques have been developed to better accommodate user-specific data distributions. Moreover, FL introduces communication and computation overhead, as frequent updates between clients and the central server can be bandwidth intensive. Optimization strategies like model compression and gradient sparsification are being explored to enhance efficiency. Finally, applying strong privacy measures often leads to reduced model accuracy or increased training time. Recent works seek to quantify and optimize this trade-off (Mohammadi et al., 2024⁶²).

Research in this field continues to evolve, with advancements in privacy-enhancing technologies like differential privacy, secure multi-party computation, and homomorphic encryption. Recent developments also include federated reinforcement learning, which extends FL to decision-making processes, and hybrid privacy models that combine FL with cryptographic techniques. Ongoing research also addresses other key challenges in FL, including data heterogeneity (non-IID data), which affects

⁶¹ <https://consensus.app/papers/secure-federations-addressing-the-security-challenges-in-ekellem/ec3525ea32a4572d989bfac6b6aaea36>

⁶² <https://consensus.app/papers/balancing-privacy-and-performance-in-federated-learning-a-mohammadi-balador/929425fda4d35153b541e68ba15b9a2c>

the convergence and performance of models. To address this, techniques such as personalized FL have been developed to better accommodate user-specific data distributions. However, FL remains vulnerable to adversarial attacks such as model poisoning or gradient inversion, where malicious participants may attempt to remain unnoticeable during transmission. Robust aggregation techniques and anomaly detection methods are actively investigated to defend against data poisoning and inference attacks. Current research trends and future directions also include:

- **Edge AI Integration:** With the growth of IoT and edge computing, FL frameworks are being tailored for low-resource devices, enabling real-time, private AI at the edge (Vishwakarma et al., 2023⁶³).
- **Regulatory-Aware AI:** Efforts are growing to embed legal compliance (e.g., GDPR) directly into AI development pipelines through privacy-preserving protocols and auditability (Shrivastava, 2024⁶⁴).
- **Hybrid Privacy Models:** Combining FL with cryptographic techniques (e.g., federated learning + homomorphic encryption) offers robust defence against both internal and external threats (Chen et al., 2024⁶⁵).

These advances underscore the importance of FL and related methodologies in promoting safe and efficient AI applications in energy systems, aligning with the growing emphasis on data privacy and regulatory compliance. As privacy regulations become stricter, the importance of Federated Learning and Privacy-Preserving AI will continue to grow because they are critical to the future of secure, efficient, and compliant AI applications across various sectors.

2.2 AI applications for the energy domain

2.2.1 Digital twins and advanced simulations

A Digital Twin (DT) is a digital representation of a physical system such as an energy asset, grid segment or building that is synchronized with real-time data in real-time to reflect its current state and forecast future behaviour. DTs go beyond static modelling by integrating live data streams, analytics, control logic and semantic reasoning to enhance understanding and decision-making across the system's lifecycle^{66 67}. On the other hand, a Digital Twin System is a hybrid system or system of systems comprising target systems, digital systems, the data relationship between them, models and data, and interfaces within that relationship. A digital twin system is made up of at least one target system and its digital twin, with synchronization and stakeholder interaction mechanisms to improve one or more performance indicators at one or more phases of the target systems' lifecycles⁶⁵.

⁶³ <https://consensus.app/papers/enabling-realtime-ai-at-the-edge-federated-learning-vishwakarma-tiwari/bb39a1025ae95a069f7cefe19762bd9d/>

⁶⁴ <https://consensus.app/papers/privacycentric-ai-navigating-the-landscape-with-shrivastava/543c58e2d4725241b262a845f8016638/>

⁶⁵ <https://consensus.app/papers/when-federated-learning-meets-privacypreserving-chen-yan/ff973a09988f527dbf21dd1d7387ef76/>

⁶⁶ <https://www.iso.org/obp/ui/en/#iso:std:iso-iec:30173:ed-1:v1:en>

⁶⁷ https://resourcecenter.ieee-pes.org/publications/technical-reports/pes_tp_tr122_itslc_052824

In the power system, DTs are playing vital role due to growing complexity, decentralization, and volatility. Deterministic traditional tools fail to replicate real-time dynamics and situational factors such as DER integration, climate change, aging infrastructure, and shifting consumption patterns. DTs offer a dynamic and adaptive environment for multi-scenario simulation, predictive maintenance, fault detection, and real-time grid optimization allowing utilities to remake operations and improve resilience^{68 66}.

AI is revolutionizing the energy sector through the integration of digital twins and advanced simulations, particularly in building energy management⁶⁹. This synergy enables real-time monitoring, predictive analytics, and optimization of energy systems, contributing to enhanced efficiency and sustainability⁷⁰. A world-renowned futurist Bernard Marr recently ranked digital twins (virtual models of physical or upcoming buildings) as one of the top trends that will reshape urban life in 2025.

Digital twins in the building context are integrating data from Building Information Modelling (BIM), Internet of Things (IoT) sensors, and AI algorithms. This integration allows for comprehensive simulation and analysis of a building's energy performance. For instance, AI-driven digital twins can model heating, ventilation, and air conditioning (HVAC) systems to identify inefficiencies and optimize energy consumption⁷¹. Incorporating AI into digital twins facilitates predictive maintenance and real-time decision-making. Studies have demonstrated that such systems can lead to significant energy savings. For example, AI applications in smart buildings have achieved up to 29% energy savings by analysing occupancy patterns and environmental conditions to optimize system operations⁷².

Advanced simulations powered by AI and digital twins enable scenario analysis and forecasting, essential for energy optimization in buildings. These simulations can assess the impact of various design choices and operational strategies on energy consumption and occupant comfort. By leveraging machine learning algorithms, simulations can predict energy demand and adjust systems proactively.

Furthermore, integrating AI with digital twins supports the development of zero-energy buildings by simulating energy flows and identifying opportunities for renewable energy integration. This approach aligns with sustainable urban development goals, promoting energy efficiency and reducing carbon footprints.

⁶⁸ <https://www.vde.com/resource/blob/2293228/3e50b3417c8777c85e9e33cb3e1e9ebc/vde-study-the-digital-twin-in-the-network-and-electricity-industry-data.pdf>

⁶⁹ Time, How AI Is Making Buildings More Energy-Efficient. Available online: <https://time.com/7201501/ai-buildings-energy-efficiency/>

⁷⁰ Smart Building International, From individual buildings to smart cities: digital twins, ai and IoT set to drive sustainability, new research reveals. Available online: <https://smartbuildingmag.com/application-stories/90524-from-individual-buildings-to-smart-cities-digital-twins%2C-ai-and-iot-set-to-drive-sustainability%2C-new-research-reveals>

⁷¹ Sghiri, A., Gallab, M., Merzouk, S., & Assoul, S. (2025). Leveraging Digital Twins for Enhancing Building Energy Efficiency: A Literature Review of Applications, Technologies, and Challenges. *Buildings*, 15(3), 498. <https://doi.org/10.3390/buildings15030498>

⁷² Yitmen, I., Almusaed, A., Hussein, M., & Almssad, A. (2025). AI-Driven Digital Twins for Enhancing Indoor Environmental Quality and Energy Efficiency in Smart Building Systems. *Buildings*, 15(7), 1030. <https://doi.org/10.3390/buildings15071030>

Some studies go in this direction. For instance, (Arsecularatne, 2024)⁷³ reviews digital twins for thermal comfort and energy efficiency; particularly, it examines various tools, technologies, and software adopted in creating digital twins, highlighting their potential in real-time monitoring and predictive maintenance. This aspect is reinforced by (Liu-Z, 2025)⁷⁴, who constructed a digital twin model for buildings to facilitate real-time monitoring and analysis of energy consumption, environmental parameters, and user behaviours. By leveraging this model, the research aims to optimize energy management strategies and enhance energy efficiency, contributing to the development of net zero energy buildings.

Another work⁷⁵ proposes an integrated approach that combines EnergyPlus simulations with AI-based optimization and decision-making models within the Mojo programming environment. The framework supports both single-criterion and multi-criteria optimization, enabling direct energy-efficient building design optimization. The study demonstrates the effectiveness of AI in enhancing building design processes for improved energy performance. Similar approach is made by (Mehraban, 2025)⁷⁶, combining EnergyPlus simulations with Feedforward Neural Networks (FNN) and Extreme Gradient Boosting (XGB) to facilitate the optimization of energy use and thermal comfort.

2.2.1.1 *Architecture and classification of DTs*

The architecture of a DT is modular and typically consists of multiple layers. It begins with the data layer, which collects and combines diverse inputs such as sensor data, SCADA, PMUs, weather information, and market signals^{77 78}. This is followed by the integration and modelling layer, where physics-based models and AI-driven simulations operate to support analysis, forecasting, and decision-making. At the top is the interaction and service layer, which provides user interfaces, visualization tools, and APIs for system control and planning. At the core of the DT lies the synchronization mechanism, which ensures continuous alignment between the digital representation and the evolving state of the physical system, maintaining accuracy and trustworthiness over time. Many utility scale DTs adopt a federated structure, coordinating multiple sub-twins such as those representing

⁷³ Arsecularatne, B., Rodrigo, N., & Chang, R. (2024). Digital Twins for Reducing Energy Consumption in Buildings: A Review. Sustainability, 16(21), 9275. <https://doi.org/10.3390/su16219275>

⁷⁴ Zhansheng Liu, Mingming Li, Weiyu Ji, Development and application of a digital twin model for Net zero energy building operation and maintenance utilizing BIM-IoT integration, Energy and Buildings, Volume 328, 2025, 115170, ISSN 0378-7788, <https://doi.org/10.1016/j.enbuild.2024.115170>

⁷⁵ Hui Liu, Zhe Du, Tingting Xue, Tao Jiang, An integrated artificial intelligence-driven approach to multi-criteria optimization of building energy efficiency and occupants' comfort: A case study, Journal of Building Engineering, Volume 103, 2025, 111944, ISSN 2352-7102, <https://doi.org/10.1016/j.jobe.2025.111944>

⁷⁶ Mohammad H. Mehraban, Samad ME Sepasgozar, Alireza Ghomimoghdam, Behrouz Zafari, AI-Enhanced Automation of Building Energy Optimization Using a Hybrid Stacked Model and Genetic Algorithms: Experiments with Seven Machine Learning Techniques and a Deep Neural Network, Results in Engineering, 2025, 104994, ISSN 2590-1230, <https://doi.org/10.1016/j.rineng.2025.104994>

⁷⁷ <https://www.iso.org/obp/ui/en/#iso:std:iso-iec:30173:ed-1:v1:en>

⁷⁸ <https://www.vde.com/resource/blob/2293228/3e50b3417c8777c85e9e33cb3e1e9ebc/vde-study-the-digital-twin-in-the-network-and-electricity-industry-data.pdf>

transformers, substations, and feeder under a supervisory DT to ensure scalability and cross-domain interoperability.

In addition, modern DTs are designed to be vendor-agnostic, secure, and cloud-enabled, leveraging open standards such as Common Information Model (CIM) and OPC Unified Architecture (OPC UA), while conforming to zero-trust cybersecurity models. These architectures integrate seamlessly with enterprise systems managed by utilities, including Energy Management Systems (EMS), Distribution Management Systems (DMS), Geographic Information Systems (GIS) etc. positioning DTs as foundational technologies in the transition toward intelligent, adaptive, and future-ready power systems. DTs can be classified into five main categories reflecting increasing levels of data integration, intelligence, and automation⁷⁹:

1. Descriptive DT: Represents structural and design information for visualization and documentation.
2. Informative DT: Adds real-time or historical data for monitoring and reporting.
3. Predictive DT: Simulates future behaviour using models and data for forecasting and maintenance.
4. Comprehensive DT (Prescriptive): Enables “what-if” scenario analysis to support optimized decision-making.
5. Autonomous DT: Operates with minimal human input using AI/ML for real-time, self-adaptive control.

These types are often combined in practice and reflect a maturity path toward more intelligent and automated power system operations.

2.2.1.2 *Digital Twin (DT) Applications in the Power Systems*

DTs are being increasingly adopted across the power system, delivering operational, analytical, and strategic benefits. In grid planning, it enables multi-scenario simulations to support infrastructure expansion, the integration of Distributed Energy Resources (DERs), and demand forecasting. They assist in decisions related to substation siting, cable undergrounding, and congestion management by representing both real-time and future system conditions. DTs also have the capacity to integrate diverse data sources to analyse the impact of consumer behaviour and climate events on network topology⁸⁰.

The following are promising current and emerging application domains for DTs in the power sector and buildings are⁸¹:

- Real-Time Operations: DTs enhance situational awareness through real-time data integration, enabling faster fault detection, dynamic contingency analysis, and real-time coordination of DERs and microgrids.
- Asset Management: DTs support condition monitoring and predictive maintenance by integrating health indicators (e.g., temperature, vibration, load cycles) with degradation models, thereby improving lifecycle decision-making.

⁷⁹ https://resourcecenter.ieee-pes.org/publications/technical-reports/pes_tp_tr122_itslc_052824

⁸⁰ <https://www.vde.com/resource/blob/2293228/3e50b3417c8777c85e9e33cb3e1e9ebc/vde-study-the-digital-twin-in-the-network-and-electricity-industry-data.pdf>

⁸¹ https://resourcecenter.ieee-pes.org/publications/technical-reports/pes_tp_tr122_itslc_052824

- **Protection and Control:** DTs facilitate real-time relay coordination, fault propagation modelling, and adaptive protection strategies, and can simulate cascading failures in systems with high levels of renewable penetration.
- **Cybersecurity:** DTs provide platforms for intrusion detection, architecture validation, and response training by replicating digital attack surfaces and cyber-physical interactions. They enable risk assessments, scenario analysis, and anomaly detection using real-time and historical data.
- **Training and Simulation:** DTs offer immersive, high-fidelity digital environments where operators can practice responses to contingencies like grid instability, extreme weather, or cyber-attacks.
- **Regulatory Compliance and Auditing:** DTs support traceable, repeatable simulations for standards verification, risk assessments, and performance benchmarking aligned with regulatory frameworks.
- **Buildings:** When applied to buildings, DTs act as dynamic, data-rich virtual replicas, synchronizing with real-time data to reflect current state, historical behaviour, and projected performance. Unlike static BIMs, DTs integrate operational data from IoT sensors, Building Automation Systems (BAS), energy meters, and environmental inputs like weather and occupancy data, enabling continuous feedback between physical and digital domains⁸². This bidirectional flow supports predictive maintenance, space utilization analytics, fault diagnostics, and energy optimization⁸³. In smart buildings, DTs enable real-time control of HVAC, lighting, and security systems, enhancing indoor environmental quality and operational efficiency⁸⁴. They also aid decarbonization by simulating retrofitting scenarios and optimizing renewable integration at the building level⁸⁵.

Collectively, these use cases affirm that DTs are not merely visualizations or static models, but intelligent systems foundational to resilient, efficient, and adaptive grid operations.

2.2.1.3 Electrolysis system's Digital Twins

Water electrolysis technologies are seen, for many reasons, as a promising way to replace conventional fossil fuels in the energy domain. Indeed, Hydrogen will accelerate the decarbonation of the energy and industrial sectors and it offers an efficient compatibility with the intermittent nature of renewable

⁸² <https://www.iso.org/obp/ui/en/#iso:std:iso-iec:30173:ed-1:v1:en>

⁸³ Volk, R., Stengel, J., & Schultmann, F. (2014). Building Information Modeling (BIM) for existing buildings — Literature review and future needs. *Automation in Construction*, 38, 109–127.
<https://doi.org/10.1016/j.autcon.2013.10.023>

⁸⁴ Boje, C., Guerriero, A., Kubicki, S., & Rezgui, Y. (2020). Towards a semantic construction digital twin: Directions for future research. *Automation in Construction*, 114, 103179. <https://doi.org/10.1016/j.autcon.2020.103179>

⁸⁵ Zhang, Y., Wang, W., Wang, Z., Gao, M., Zhu, L., & Song, J. (2021). Green building design based on solar energy utilization: Take a kindergarten competition design as an example. *Energy Reports*, 7, 1297–1307.
<https://doi.org/10.1016/j.egyr.2021.09.134>

energy such as solar and wind. These applications require multi-megawatt (MW) scale systems, with guaranteed service continuity and operation time for various conditions (Iyer et al. 2024⁸⁶).

Real testing is essential to prove the relevance of these technologies but has limits when the high-power scale and long-time operation topics are considered (Aicart et al. 2023⁸⁷). To treat these subjects, the system modeling approach is seen as an interesting tool.

The development of electrolysis system's Digital Twins has become a key tool for many applications, such as the design and sizing at the early phase of projects or to develop the system's control (Cai et al. 2014⁸⁸). In the literature, Solid Oxide Electrolysis (SOE) models are classified by scale (Iyer et al. 2024) (Li et al. 2021⁸⁹). Stack-level models describe voltage, flow and temperature distribution for several cells or complete stack components with various levels of detail for geometry. At system-level, stacks and the Balance of Plant (BoP) are represented in a simple manner; interactions between stacks are not represented.

Some studies relate the use of machine learning techniques to improve the development of Anion-Exchange Membrane electrolysis (AEMEL) (Günay et al. 2023⁹⁰) and Proton Exchange Membrane electrolysis (PEMEL) (Ding et al. 2023⁹¹). They are focused on the improvement at the cell and stack scale and more precisely at the materials level. For the Solid Oxide Cell Electrolysis (SOEC) technology, no such advanced methods are involved in the literature. Botta et al.⁹² or Pohjoranta et al.⁹³ present the use of model predictive control (MPC) applied to the thermal management of the stacks. System models are developed for optimizing global system efficiency with investigation on design and

⁸⁶ Iyer, Siddharth et al. (2024-06). "Review of experimental and modelling investigations for solid oxide electrolysis technology". *International Journal of Hydrogen Energy* 72, pp. 537–558. DOI: 10.1016/j.ijhydene.2024.05.312.

⁸⁷ J. Aicart, A. Surrey, L. Champelovier, K. Henault, C. Geipel, O. Posdziech, J. Mougine, *Fuel Cells* 2023, 23, 463. <https://doi.org/10.1002/fuce.202300028>

⁸⁸ Cai, Qiong & Adjiman, Claire & Brandon, Nigel. (2014). Optimal control strategies for hydrogen production when coupling solid oxide electrolyzers with intermittent renewable energies. *Journal of Power Sources*. 268. 212–224.

⁸⁹ Z. Li, H. Zhang, H. Xu, and J. Xuan, "Advancing the multiscale understanding on solid oxide electrolysis cells via modelling approaches: A review," *Renew. Sustain. Energy Rev.*, vol. 141, p. 110863, May 2021

⁹⁰ M. Erdem Günay, N. Alper Tapan, Analysis of PEM and AEM electrolysis by neural network pattern recognition, association rule mining and LIME, *Energy and AI*, Volume 13, 2023, 100254, ISSN 2666-5468, <https://doi.org/10.1016/j.egyai.2023.100254>.

⁹¹ Rui Ding, Yawen Chen, Zhiyan Rui, Kang Hua, Yongkang Wu, Xiaoke Li, Xiao Duan, Jia Li, Xuebin Wang, Jianguo Liu, Machine learning utilized for the development of proton exchange membrane electrolyzers, *Journal of Power Sources*, Volume 556, 2023, 232389, ISSN 0378-7753, <https://doi.org/10.1016/j.jpowsour.2022.232389>.

⁹² G. Botta, M. Romeo, A. Fernandes, S. Trabucchi, P.V. Aravind, Dynamic modeling of reversible solid oxide cell stack and control strategy development, *Energy Conversion and Management*, Volume 185, 2019, Pages 636-653, ISSN 0196-8904, <https://doi.org/10.1016/j.enconman.2019.01.082>.

⁹³ Antti Pohjoranta, Matias Halinen, Jari Pennanen, Jari Kiviaho, Model predictive control of the solid oxide fuel cell stack temperature with models based on experimental data, *Journal of Power Sources*, Volume 277, 2015, Pages 239-250, ISSN 0378-7753, <https://doi.org/10.1016/j.jpowsour.2014.11.126>.

operating strategies with tests of control algorithm or for techno-economic assessment. These purposes do not require detailed thermal behaviour of the hotbox and the stacks. Between stack and system level, a level can be considered for multi-stacks module. This level allows to investigate the impact of stacks layout on thermal and fluidic aspect or to developed specific thermal management of the module for example. Few studies represent multi-stacks module behaviour. They are mainly performed with 3D CFD model as in Chi et al.⁹⁴ where a geometry homogenization approach is used on stack level to reduce the calculation time at 5% computational time of a full geometry model, with prediction error less than 1°C. Yang et al.⁹⁵ described a 3D MSM model using Distributed Resistance Analogy (DRA) methods. This study investigates the effect of module architecture -numbers of stacks, cells and connection configuration- under operation in electrolysis or fuel cell mode. Regarding long-term simulation, Giridhar et al.⁹⁶ perform as analysis up to 20000 operating hours considering long-term chemical degradation in a 2D planar SOE cell model. Many simplification hypotheses regarding to the thermal maps in the stacks and furthermore with the thermal exchange with the environment are considered.

2.2.2 AI for Smart Grid Management

Artificial intelligence (AI) has become a strategic tool for the modernisation of smart grids, particularly in its ability to automate complex processes and adapt to the dynamic nature of the electricity system. In this context, machine learning (ML) has proven to be particularly effective in tasks such as load forecasting, renewable generation estimation, anomaly detection and demand management.

Models like LSTM networks, transformers, and convolutional neural networks enable more accurate predictions than traditional methods, reducing overproduction and energy losses.

Among the most used techniques are regression models, decision trees and ensemble methods such as random forests or XGBoost, which allow flexible modelling of historical patterns of electricity consumption and production. These techniques have been applied in both centralised environments and distributed systems, improving operational efficiency by anticipating grid behaviour and reducing unnecessary generation (Ortiz-Torres et al., 2024⁹⁷).

In recent years, deep learning (DL) has outperformed traditional approaches by incorporating models capable of learning hierarchical representations and capturing complex temporal dynamics. Recurrent

⁹⁴ Y. Chi, J. Lin, P. Li, and Y. Song, "Investigating the performance of a solid oxide electrolyzer multi-stack module with a multiphysics homogenized model," *J. Power Sources*, vol. 594, p. 234019, Feb. 2024, doi: 10.1016/j.jpowsour.2023.234019.

⁹⁵ C. Yang et al., "Modeling and analysis of the multiphysics transport parameters of a kilowatt-class multistack module of reversible solid oxide cells," *Appl. Therm. Eng.*, vol. 235, p. 121373, Nov. 2023, doi: 10.1016/j.applthermaleng.2023.121373.

⁹⁶ N. V. Giridhar, D. A. Allan, M. Li, S. E. Zitney, L. T. Biegler, and D. Bhattacharyya, "Optimal operation of solid-oxide electrolysis cells considering long-term chemical degradation," *Energy Convers. Manag.*, vol. 319, p. 118950, Nov. 2024, doi: 10.1016/j.enconman.2024.118950.

⁹⁷ Ortiz-Torres, L. F., Gómez-Luna, E., & Sáenz, E. M. (2024). Estudio del uso y contribución de la inteligencia artificial para la operación en redes eléctricas. *Revista UIS Ingenierías*, 23(2), 31-46.

models such as LSTM (Long Short-Term Memory) (Ahmed et al., 2023⁹⁸) and GRU (Gated Recurrent Units) networks are widely used for short-term electric load prediction, surpassing in accuracy models such as SVM or ARIMA (Shiri et al., 2023⁹⁹). Furthermore, it has been observed that the combination of recurrent networks with convolutional networks and attention mechanisms, as in hybrid LSTM-CNN-Attention models, further improves performance in multivariate forecasting tasks. Such architectures allow not only to capture temporal relationships, but also to automatically identify the most relevant variables in each operational context (Papastefanopoulos et al., 2023¹⁰⁰).

A fast-growing area is the application of AI in fault detection and localisation, where deep learning-based classification models have demonstrated response capabilities in the order of milliseconds. This is essential in systems where the speed of reaction to disturbances can prevent outages or damage to critical infrastructure (Najafzadeh, Masoud, et al., 2024¹⁰¹). Furthermore, anomaly detection algorithms such as autoencoders, isolation forests and residual prediction models are used to identify potential power fraud, technical losses or errors in smart meters, even under conditions of high uncertainty (Javaid, Nadeem, et al., 2022¹⁰²).

2.2.2.1 *Machine Learning Approaches in Smart Grids*

Machine learning has become a powerful tool for enhancing smart grid operations and energy efficiency. Traditional ML algorithms (such as regression models, support vector machines, decision trees, and ensemble methods) have been widely used for tasks like load forecasting, anomaly detection, and energy usage optimization. ML-based models are used to predict electricity demand and renewable generation (e.g. wind power output) more accurately, and to optimize how energy is dispatched or stored in response. (Kiasari, M., Ghaffari, M., & Aly, H. H., 2024¹⁰³). Overall, supervised learning models that learn from historical data have proven effective to reduce energy wastage and enhance efficiency in smart grids. Utilities report that ML-driven load forecasts and demand response recommendations

⁹⁸ Ahmed, S. F., Alam, M. S. B., Hassan, M., Rozbu, M. R., Ishtiaq, T., Rafa, N., ... & Gandomi, A. H. (2023). Deep learning modelling techniques: current progress, applications, advantages, and challenges. *Artificial Intelligence Review*, 56(11), 13521-13617

⁹⁹ Shiri, F. M., Perumal, T., Mustapha, N., & Mohamed, R. (2023). A comprehensive overview and comparative analysis on deep learning models: CNN, RNN, LSTM, GRU. *arXiv preprint arXiv:2305.17473*.

¹⁰⁰ Papastefanopoulos, V., Linardatos, P., Panagiotakopoulos, T., & Kotsiantis, S. (2023). Multivariate time-series forecasting: A review of deep learning methods in internet of things applications to smart cities. *Smart Cities*, 6(5), 2519-2552.

¹⁰¹ Najafzadeh, M., Pouladi, J., Daghigh, A., Beiza, J., & Abedinzade, T. (2024). Fault detection, classification and localization along the power grid line using optimized machine learning algorithms. *International Journal of Computational Intelligence Systems*, 17(1), 49.

¹⁰² Javaid, N., Qasim, U., Yahaya, A. S., Alkhamash, E. H., & Hadjouni, M. (2022). Non-technical losses detection using autoencoder and bidirectional gated recurrent unit to secure smart grids. *IEEE Access*, 10, 56863-56875.

¹⁰³ Kiasari, M., Ghaffari, M., & Aly, H. H. (2024). A Comprehensive Review of the Current Status of Smart Grid Technologies for Renewable Energies Integration and Future Trends: The Role of Machine Learning and Energy Storage Systems. *Energies*, 17(16), 4128. <https://doi.org/10.3390/en17164128>

can improve energy efficiency and cut down unnecessary generation. (Balamurugan, M., Narayanan, K., Raghu, N., Arjun Kumar, G. B., & Trupti, V. N., 2025¹⁰⁴)

One study found that a regression-tree model could successfully detect suspicious data injections in a distributed solar generation system, outperforming earlier methods like SVM or ARIMA in identifying energy theft attempts (Guato Burgos, M. F., Morato, J., & Vizcaino Imacaña, F. P. 2024¹⁰⁵).

2.2.2.2 Deep Learning Advances in Smart Grid Management

Recurrent deep models like Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRUs) have become standard for short-term load forecasting. They capture temporal dependencies far better than shallow models. Recent studies have further enhanced these with attention mechanisms and convolutional layers (Wen et al. 2024¹⁰⁶) propose a hybrid deep model combining a GRU (to capture long-term temporal dependencies) with a Temporal Convolutional Network (TCN) for local pattern extraction, plus an attention mechanism to focus on relevant features. On multiple public datasets, their model outperformed baseline methods in prediction accuracy and efficiency.

Researchers report that well-trained deep models can automatically recognize and locate faults within milliseconds of occurrence (Alhanaf et al. 2023¹⁰⁷).

2.2.3 Optimization of Energy Demand and Resources

Artificial Intelligence (AI) has emerged as a transformative force in optimizing energy demand and resource management within buildings, addressing the pressing need for energy efficiency and sustainability. Machine Learning (ML) techniques, including deep learning, reinforcement learning, and unsupervised learning, have been instrumental in enhancing building energy efficiency (Fayyaz et al., 2024¹⁰⁸). Deep learning models effectively capture complex nonlinear relationships, enabling accurate predictions of building performance and energy consumption (Wu et al., 2023¹⁰⁹). Reinforcement learning dynamically adjusts building operational parameters through interactive learning with the environment, achieving optimal energy control.

AI-driven systems have been successfully implemented in real-world scenarios. For instance, BrainBox AI's ARIA platform automates HVAC optimization in large commercial buildings by continuously

¹⁰⁴ Balamurugan, M., Narayanan, K., Raghu, N., Arjun Kumar, G. B., & Trupti, V. N. (2025). Role of artificial intelligence in smart grid - a mini review. *Frontiers in artificial intelligence*, 8, 1551661.

<https://doi.org/10.3389/frai.2025.1551661>

¹⁰⁵ Guato Burgos, M. F., Morato, J., & Vizcaino Imacaña, F. P. (2024). A Review of Smart Grid Anomaly Detection Approaches Pertaining to Artificial Intelligence. *Applied Sciences*, 14(3), 1194.

<https://doi.org/10.3390/app14031194>

¹⁰⁶ Wen, X., Liao, J., Niu, Q. et al. Deep learning-driven hybrid model for short-term load forecasting and smart grid information management. *Sci Rep* 14, 13720 (2024). <https://doi.org/10.1038/s41598-024-63262-x>

¹⁰⁷ Alhanaf, A. S., Balik, H. H., & Farsadi, M. (2023). Intelligent Fault Detection and Classification Schemes for Smart Grids Based on Deep Neural Networks. *Energies*, 16(22), 7680. <https://doi.org/10.3390/en16227680>

¹⁰⁸ Fayyaz, S., Qureshi, R. S., Rehman, A. U., & Murtaza, M. A. (2024). A review of machine learning and AI-based approaches for building energy management. *Energy Reports*, 10, 161–180.

<https://doi.org/10.1016/j.egyr.2024.01.050>

¹⁰⁹ Wu, Y., Wu, D., & Wang, R. (2023). A review of deep reinforcement learning methods for building energy efficiency. *Buildings*, 15(7), 994.

monitoring data points such as humidity and ventilation rates, forecasting future conditions, and taking corrective actions to enhance efficiency (BrainBox AI, 2024¹¹⁰). This system has been deployed in over 14,000 buildings across more than 20 countries, claiming to reduce energy costs by 25% and significantly decrease greenhouse gas emissions.

Recent studies have increasingly focused on how AI can support both energy savings and occupant comfort. Deep learning models, such as recurrent neural networks (RNNs), are employed to predict indoor climate variables and adjust system operations accordingly. For example, an end-to-end deep meta-modelling approach uses RNNs trained on building simulation data to replicate thermal behaviour and indoor air quality under various operational scenarios (Bouchiha et al., 2021¹¹¹). These models are then calibrated using sensor data from real buildings and optimized using multi-objective algorithms that minimize energy consumption while preserving thermal comfort and CO₂ thresholds. This allows for adaptive control strategies that respond in real time to fluctuating internal and external conditions, offering substantial energy savings without compromising user well-being.

In addition to real-time operational optimization, AI is increasingly being applied to support decision-making in building refurbishment. By analysing large datasets from building stock, AI can help identify cost-effective renovation strategies that prioritize energy efficiency improvements (González-Burgos & Rojo, 2024¹¹²). These tools consider building geometry, materials, historical consumption patterns, and climate data to recommend tailored retrofitting actions. Such predictive modelling supports not only the technical feasibility of upgrades but also enhances their economic and environmental impact, contributing to decarbonization goals in the building sector.

Optimisation of energy demand and resources has become a fundamental pillar for the development of sustainable and efficient energy systems. This optimisation involves the integration of various technologies and approaches, from accurate demand forecasting to the effective management of energy resources, with the aim of minimising costs and maximising operational efficiency (Arévalo et al., 2024¹¹³). Considering also energy demand in buildings, this optimisation encompasses not only the management of energy in electricity grids, but also the adaptation of energy consumption systems to the needs of urban and residential spaces (Hernández et al., 2024¹¹⁴).

In general terms, demand-side optimisation focuses on the ability to anticipate and adapt energy consumption to variations in supply and demand (Nebey, A.H., 2024¹¹⁵). This is achieved using

¹¹⁰ BrainBox AI. (n.d.). Using artificial intelligence to optimize the flow of energy through the built environment. Retrieved from <https://brainboxai.com/en/white-papers/using-artificial-intelligence-to-optimize-the-flow-of-energy-through-the-built-environment>

¹¹¹ Bouchiha, H., Bensaid, M., & Fouchal, H. (2021). Deep learning-based meta-modeling and optimization for building energy and comfort performance. *arXiv*. <https://arxiv.org/abs/2105.02814>

¹¹² González-Burgos, C. A., & Rojo, J. C. (2024). Artificial Intelligence in sustainable building renovation: A systematic review. *Sustainability*, 16(9), 3627. <https://doi.org/10.3390/su16093627>

¹¹³ Arévalo, P., Ochoa-Correa, D., & Villa-Ávila, E. (2024). Optimizing microgrid operation: Integration of emerging technologies and artificial intelligence for energy efficiency. *Electronics*, 13(18), 3754.

¹¹⁴ Hernández, J. L., de Miguel, I., Vélez, F., & Vasallo, A. (2024). Challenges and opportunities in European smart buildings energy management: A critical review. *Renewable and Sustainable Energy Reviews*, 199, 114472.

¹¹⁵ Nebey, A. H. (2024). Recent advancement in demand side energy management system for optimal energy utilization. *Energy Reports*, 11, 5422-5435.

advanced predictive models and optimisation algorithms, such as linear programming (LP) models, evolutionary algorithms and machine learning models. These approaches allow not only the reduction of energy consumption based on actual demand, but also the more efficient integration of renewable energy sources, such as solar or wind, which are inherently intermittent (Kafakthong, Natdanai, 2024¹¹⁶).

One of the areas where demand-side optimisation has a significant impact is in heating, ventilation and air conditioning (HVAC) systems, both in smart buildings and industrial facilities (Köse, Ahmet et al., 2024¹¹⁷). Predictive optimisation algorithms, which use artificial intelligence-based techniques to predict future energy needs, enable the dynamic control of HVAC systems, adjusting temperature and humidity in real time to improve energy efficiency without compromising user comfort. In this way, energy demand is aligned with anticipated behaviour, reducing energy waste, which is particularly relevant in buildings that have varying energy demands throughout the day (Hanafi et al., 2024¹¹⁸).

It is important to note that resource optimisation extends beyond internal demand management to include interaction with smart grids. Here, buildings play a key role through demand response, where consumers adjust their demand in response to electricity price signals or system needs. This type of response is essential to balance fluctuations between supply and demand in real time, improving the stability of the electricity system and helping to avoid demand peaks that can lead to blackouts or the use of non-renewable energy sources (Khan, Muhammad Adnan, et al., 2022¹¹⁹).

Another crucial dimension in the optimisation of energy demand lies in the heterogeneity of users and usage profiles. Industrial facilities, residential consumers, and public administration buildings exhibit vastly different load patterns, operational constraints, and flexibility potential. AI-based optimisation frameworks can accommodate this diversity by tailoring control strategies to each user category while coordinating their operation within local energy communities or district-level energy management systems. For example, industry may offer large but rigid loads, public buildings can contribute through scheduled HVAC or lighting control, and households may offer responsive flexibility via smart appliances or EV charging. Coordinating such heterogeneous actors enables system-wide optimisation while meeting sector-specific priorities such as cost, comfort, or regulatory compliance (Zatti et al., 2021¹²⁰ and Mazzarino et al., 2022¹²¹).

¹¹⁶ Kafakthong, N. (2024). Prediction of optimal binding constraints of linear programming problem using deep learning.

¹¹⁷ Köse, A., Sukhanov, I., Belikov, J., Tepljakov, A., & Petlenkov, E. (2024, June). Demand Side Management for Commercial Buildings: Multiple Heating Systems Based Optimization in Real-Time. In *2024 IEEE International Conference on Engineering, Technology, and Innovation (ICE/ITMC)* (pp. 1-8). IEEE.

¹¹⁸ Khan, M. A., Saleh, A. M., Waseem, M., & Sajjad, I. A. (2022). Artificial intelligence enabled demand response: Prospects and challenges in smart grid environment. *Ieee Access*, 11, 1477-1505.

¹¹⁹ Khan, M. A., Saleh, A. M., Waseem, M., & Sajjad, I. A. (2022). Artificial intelligence enabled demand response: Prospects and challenges in smart grid environment. *Ieee Access*, 11, 1477-1505.

¹²⁰ Zatti, M., Moncecchi, M., Gabba, M., Chiesa, A., Bovera, F., & Merlo, M. (2021). Energy Communities Design Optimization in the Italian Framework. *Applied Sciences*, 11(11), 5218. <https://doi.org/10.3390/app11115218>

¹²¹ P. R. Mazzarino, M. Capone, E. Guelpa, V. Verda, L. Bottaccioli and E. Patti, "A Multi-Agent framework to evaluate energy flexibility in District Heating networks," *2022 IEEE International Conference on Environment and*

In addition, the use of distributed energy resources (DERs), such as solar panels and energy storage systems, facilitates the optimisation of resources within buildings and grids. By integrating these resources with hybrid optimisation algorithms that combine machine learning and mathematical optimisation, the use of renewable energy can be maximised, reducing dependence on the grid and improving the energy autonomy of buildings. This integration also allows buildings to optimise their energy consumption and storage, generating strategies that can be transferred to real environments with little retraining (Twaïsan, Kumail, and Necaattin Barışçı, 2022¹²²).

Furthermore, multi-objective optimisation is becoming increasingly relevant for energy systems that must balance competing goals such as cost efficiency, carbon reduction, and user comfort. In hybrid systems that integrate renewables, batteries, and grid power, artificial intelligence techniques—such as reinforcement learning, genetic algorithms, and swarm intelligence—are used to generate smart scheduling strategies. These algorithms can forecast load, weather, and market conditions to dynamically decide when to store, consume, or sell energy. For instance, in smart buildings and local energy communities, AI-based control systems can optimise for both economic savings and environmental performance, adapting to real-time constraints and evolving user preferences (Li et al., 2020¹²³; Zhang et al., 2021¹²⁴). Such flexible, goal-aware energy management is essential to ensure system-wide resilience and sustainability. Finally, advanced communication through the Internet of Things (IoT) and integrated energy management systems plays a key role in optimising demand and resources. Real-time data collection and analysis provide crucial information to adjust energy system behaviour efficiently, improving not only consumption efficiency, but also the reliability and sustainability of energy systems in general (Saleem, M. Usman, et al., 2023¹²⁵).

2.2.4 Predictive Maintenance and AI-Based Diagnostics

Predictive maintenance (PdM) has emerged as a cornerstone of modern asset management strategies, offering a paradigm shift from traditional time-based maintenance to a more intelligent, condition-based approach. Recent progress in PdM is largely attributed to the development of low-cost, high-resolution sensors and real-time condition monitoring systems, which have enabled the collection of large-scale operational data. These advancements, combined with sophisticated algorithms and the valuable insights of human experts, have significantly improved the precision and effectiveness of predictive maintenance strategies (Ucar et al., 2024¹²⁶). Drawing on recent advances in artificial intelligence (AI),

Electrical Engineering and 2022 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe), Prague, Czech Republic, 2022, pp. 1-6, doi: 10.1109/EEEIC/ICPEurope54979.2022.9854622.

¹²² Twaïsan, K., & Barışçı, N. (2022). Integrated distributed energy resources (DER) and microgrids: Modeling and optimization of DERs. *Electronics*, 11(18), 2816.

¹²³ Li, K., Zhang, T., & Liu, H. (2020). *Deep reinforcement learning for multi-objective energy management in hybrid systems*. *Applied Energy*, 275, 115398. <https://doi.org/10.1016/j.apenergy.2020.115398>

¹²⁴ Zhang, Y., Sun, M., & Zhou, Q. (2021). *Multi-objective optimisation for smart energy systems: A comprehensive review*. *Renewable and Sustainable Energy Reviews*, 135, 110206. <https://doi.org/10.1016/j.rser.2020.110206>

¹²⁵ Saleem, M. U., Shakir, M., Usman, M. R., Bajwa, M. H. T., Shabbir, N., Shams Ghahfarokhi, P., & Daniel, K. (2023). Integrating smart energy management system with internet of things and cloud computing for efficient demand side management in smart grids. *Energies*, 16(12), 4835.

¹²⁶ Ucar, A.; Karakose, M.; Kırımçı, N. Artificial Intelligence for Predictive Maintenance Applications: Key Components, Trustworthiness, and Future Trends. *Appl. Sci.* **2024**, *14*, 898. <https://doi.org/10.3390/app14020898>

particularly machine learning and deep learning, PdM enables early detection of faults and degradation patterns in critical infrastructures. By analysing sensor data, operational logs, and equipment usage patterns, AI-driven PdM facilitates the anticipation of failures before they occur, thereby reducing unplanned downtime, extending equipment lifespan, and minimizing operational and maintenance costs. Unlike conventional preventive maintenance, which relies on fixed schedules and statistical lifetime estimates, predictive maintenance dynamically adapts to the actual state of components and systems. AI models trained on historical and/or real-time data can learn complex failure signatures and subtle precursors to failure, often imperceptible to human operators or rule-based systems. These models outperform traditional rule-based or threshold-based diagnostics, particularly in dynamic and harsh operating conditions (Susto et al., 2015¹²⁷).

Within the energy sector, PdM has seen growing adoption across critical infrastructure, including transformers, wind turbine gearboxes, battery storage units, and grid assets. Embedded AI models enable anomaly detection directly at the edge, enabling real-time inference that reduces latency and improves decision-making, an essential feature for managing geographically distributed assets (Zhang et al., 2019¹²⁸). Moreover, the integration of edge computing further strengthens system resilience by minimizing dependence on continuous cloud connectivity, ensuring robust maintenance operations even under constrained network conditions. Furthermore, the deployment of digital twins is transforming PdM strategies by enabling the simulation of operational scenarios and the early prediction of component failures.

Despite these advancements, challenges persist, including ensuring data quality, integrating AI technologies into existing systems, and addressing ethical considerations related to AI deployment. Ongoing research continues to focus on enhancing the transparency and explainability of AI models to foster trust and facilitate broader adoption in PdM applications (Ucar et al., 2024¹²⁹).

2.2.4.1 *Health management*

Each stack has a unique degradation process due to its manufacturing conditions, its operating conditions but also to its operating history. So, to have a best performance loss, management indicators that reflect it are needed. That's what is called **health management**. Health management is based on data received from sensors. Data are analysed to give an indicator that can help in:

- condition assessment: giving an indication of the state of life or detecting or localizing fault,
- diagnosis: identifying and determining the cause of the fault,
- prognosis: trying to estimate the end of life to help for maintenance decision for example,

To diagnose a stack or prognosticate its failure the main indicators available in literature are based on the EIS (Electrochemical Impedance Spectroscopy) or current-voltage (IV curve or polarization curve) curves. These ones are used to characterize physical feature: the ohmic, polarization and total

¹²⁷ G. A. Susto, A. Schirru, S. Pampuri, S. McLoone, and A. Beghi, "Machine Learning for Predictive Maintenance: A Multiple Classifier Approach," IEEE Transactions on Industrial Informatics, vol. 11, no. 3, pp. 812–820, June 2015. <https://doi.org/10.1109/TII.2014.2349359>

¹²⁸ W. Zhang, D. Yang and H. Wang, "Data-Driven Methods for Predictive Maintenance of Industrial Equipment: A Survey," in IEEE Systems Journal, vol. 13, no. 3, pp. 2213–2227, Sept. 2019, doi: 10.1109/JSYST.2019.2905565

¹²⁹ Ucar, A.; Karakose, M.; Kırımca, N. Artificial Intelligence for Predictive Maintenance Applications: Key Components, Trustworthiness, and Future Trends. Appl. Sci. **2024**, *14*, 898. <https://doi.org/10.3390/app14020898>

resistance, also called ASR (Area specific resistance), the capacity of the double layer, load diffusion impedance, ... These tools are usually performed in constant operating conditions: temperature, gas pressure, stoichiometry rates and realized at regular time intervals for different sets of operating conditions. This set of curves can be used to determine the loss of performance of stack and its evolution by way of comparison^{130 131}.

From these curves data-based methods propose to use features of these ones to detect a fault. The principle is to build a model of the features based on the collected data and compare it to the measurements. The difference is called residues. According to these residues we can determine if we are in normal situation or if there is a fault^{132 133 134 135}. For instance, Fouquet and al.¹³⁶ model the fuel cell as a RC circuit to identify the parameter resistant thanks to EIS measurement. The ohmic system (polarization, membrane and diffusion resistances) helps to determine the fuel cell state of health. According to the location in the 3D space formed by these three resistances, the state of flooding or drying or "normal" can be deduced. These methods are often linked to a diagnosis. The model leads to cluster that are relevant to the state of the stack. For instance, Ghorbani and al.¹³⁷ study evaluated Naïve Bayes, Logistic Regression, KNN, and ANN to classify the cell into "starved" and "not starved".

¹³⁰ R Onanena, L Oukhellou, E Côme, D Candusso, D Hissel, P Aknin, Fault-diagnosis of PEM fuel cells using electrochemical spectroscopy impedance, IFAC Proceedings Volumes, Volume 45, Issue 21.
<https://doi.org/10.3182/20120902-4-FR-2032.00114>.

¹³¹ [Daniel B. Drasbæk](#), [Peter Blennow](#), [Thomas Heiredal-Clausen](#), [Jeppe Rass-Hansen](#), [Giovanni Perin](#), [Jens V. T. Høgh](#), [Anne Hauch](#), Exploring electrochemical impedance spectroscopy to identify and quantify degradation in commercial solid oxide electrolysis stacks. <https://doi.org/10.1002/fuce.202300023>.

¹³² R Onanena, L Oukhellou, E Côme, D Candusso, D Hissel, P Aknin, Fault-diagnosis of PEM fuel cells using electrochemical spectroscopy impedance, IFAC Proceedings Volumes, Volume 45, Issue 21.
<https://doi.org/10.3182/20120902-4-FR-2032.00114>.

¹³³ D. Hissel, M.C. Pera, Diagnostic & health management of fuel cell systems: Issues and solutions, Annual Reviews in Control, Volume 42, 2016. <https://doi.org/10.1016/j.arcontrol.2016.09.005>.

¹³⁴ Pierpaolo Polverino, Marco Sorrentino, Cesare Pianese, A model-based diagnostic technique to enhance faults isolability in Solid Oxide Fuel Cell systems, Applied Energy, Volume 204, 2017.
<https://doi.org/10.1016/j.apenergy.2017.05.069>.

¹³⁵ Pahon, E., Yousfi-Steiner, N., Jemei, S., Hissel, D., & Moçoteguy, P. (2017). A Non-Intrusive Signal-Based Method for a Proton Exchange Membrane Fuel Cell Fault Diagnosis. *Fuel Cells*, 17(2), 238-246.
<https://doi.org/10.1002/fuce.201600070>.

¹³⁶ Fouquet, N., Doulet, C., Nouillant, C. J. J., Dauphin-Tanguy, G., & Ould-Bouamama, B. (2006). Model based PEM fuel cell state-of-health monitoring via ac impedance measurements. *Journal of Power Sources*, 159(2), 905-913.
<https://doi.org/10.1016/j.jpowsour.2005.11.035>.

¹³⁷ Ghorbani, M., Boley, M., Nakashima, P. N. H., & Biribilis, N. (2024). An active machine learning approach for optimal design of magnesium alloys using Bayesian optimisation. *Scientific Reports*, 14(1), 8299.
<https://doi.org/10.1038/s41598-024-59100-9>.

Another method to do prognosis is to build a prediction model of the end of life thanks to a variable revealing the state of health. These models are often based on a Kalman approach^{138 139 140}. Bressel and al.¹⁴¹ choose features of the IV curves to detect the state of health of the fuel cell. Zhang and al.¹⁴² studied the degradation rate based on the effective electro-chemical surface area¹⁴³.

At each time step the state of health is compared to a threshold thanks to a metric¹⁴⁴ to determine the time of failure. The metrics chosen can be Norm 1 or 2. Other methods exist to determine the failure time, for example Zhang and al.¹⁴⁵ propose a probabilistic method.

The difficulty is to obtain a good prediction model. Most of the time, these models are precise in a short temporal horizon and need to be updated regularly with ageing. A change of mode will also considerably change the result.

Over the life of the battery, operating conditions change depending on the management strategy and its aging. Therefore, running an EIS or IV curve will change the current usage. This is why it is important to have an online diagnostic method. To over pass the problem of the operating conditions variability to do diagnosis Nitsche and al.¹⁴⁶ or Jemei and al.¹⁴⁷ proposed a method based on neuronal modelling to establish a diagnosis of the fuel cell. The aim is to plot a IV curve in situation of driving. Nitsche and

¹³⁸ Bressel, M., Hilaiet, M., Hissel, D., & Bouamama, B. O. (2016). Extended Kalman filter for prognostic of proton exchange membrane fuel cell. *Applied Energy*, 164, 220-227. <https://doi.org/10.1016/j.apenergy.2015.11.071>.

¹³⁹ Zhang, X., & Pisu, P. (2014). An unscented Kalman filter based on-line diagnostic approach for PEM fuel cell flooding. *International Journal of Prognostics and Health Management*, 5(1). <https://doi.org/10.36001/ijphm.2014.v5i1.2204>.

¹⁴⁰ Amedi, H., Pishvaie, M. R., & Golzari, A. (2014). Comparative Studies on Performance of Variants of Kalman Filters for Controlling a Sofc Unit.

¹⁴¹ Bressel, M., Hilaiet, M., Hissel, D., & Bouamama, B. O. (2016). Extended Kalman filter for prognostic of proton exchange membrane fuel cell. *Applied Energy*, 164, 220-227. <https://doi.org/10.1016/j.apenergy.2015.11.071>.

¹⁴² Zhang, X., & Pisu, P. (2014). An unscented Kalman filter based on-line diagnostic approach for PEM fuel cell flooding. *International Journal of Prognostics and Health Management*, 5(1). <https://doi.org/10.36001/ijphm.2014.v5i1.2204>.

¹⁴³ Amedi, H., Pishvaie, M. R., & Golzari, A. (2014). Comparative Studies on Performance of Variants of Kalman Filters for Controlling a Sofc Unit.

¹⁴⁴ Saha B, Goebel K, Christophersen J. Comparison of prognostic algorithms for estimating remaining useful life of batteries. *Transactions of the Institute of Measurement and Control*. 2009;31(3-4):293-308. doi:[10.1177/0142331208092030](https://doi.org/10.1177/0142331208092030)

¹⁴⁵ Zhang, X., & Pisu, P. (2014). An unscented Kalman filter based on-line diagnostic approach for PEM fuel cell flooding. *International Journal of Prognostics and Health Management*, 5(1). <https://doi.org/10.36001/ijphm.2014.v5i1.2204>.

¹⁴⁶ Nitsche, C., Schroedl, S., & Weiss, W. (2004, June). Onboard diagnostics concept for fuel cell vehicles using adaptive modelling. In *IEEE Intelligent Vehicles Symposium, 2004* (pp. 127-132). IEEE. DOI: [10.1109/IVS.2004.1336368](https://doi.org/10.1109/IVS.2004.1336368)

¹⁴⁷ Jemei, S. (2004). *Neural Network Model of a PEM Fuel Cell Power Generator* (No. FRNC-TH--11995). Universite de Technologie de Belfort Montbeliard et Universite de Franche Comte, Laboratoire de recherche en Electronique Electrotechnique et Systemes-L2ES, Unite Mixte UTBM-UFC-Inrets (France).

al.¹⁴⁸ propose to create a neuronal network to model the voltage considering the variations of the current density, temperature, air and hydrogen pressure and the air flow. Once the model is learned they use this model with pre-specified reference inputs. It will estimate the characteristic curve as it should be observed under constant test bench conditions. They can then compare this curve to a reference and hence to diagnose short term failures or errors according to fuel cell degradation.

Regarding the SOEC, Subotić et al.¹⁴⁹ developed ANN model with the aim of determining the polarization curve and cell impedance as functions of the operating temperature, current density and various fuel compositions. Xu and al. proposed a hybrid method combining an MPS and deep learning to predict the key characteristics of SOFCs operating with a complex fuel composition. Ghorbani and al.¹⁵⁰ developed a virtual hydrogen sensor to detect fuel starvation based on ANN. Zhang and al.¹⁵¹ built ELM (Extreme Learning Machine) to predict the characteristics of the SOEC under various fuel compositions within the varied range of training conditions.

2.3 HPC and Big Data for AI-Driven Energy Solutions

2.3.1 High-Performance Computing (HPC)

2.3.1.1 High-Performance Computing in the Energy Sector

HPC is a cornerstone technology in the energy sector, enabling organizations to process vast datasets, run complex simulations, and accelerate innovation across oil & gas, renewables, and utilities. HPC consolidates immense computation power, allowing for real-time analytics, predictive modeling, and optimization that were previously unattainable with traditional IT systems.

*Key Applications of HPC in Energy*¹⁵²

- Simulation and Modeling:
HPC enables advanced simulations for wind farm layouts, reservoir modeling, and combustion processes, leading to optimized designs, increased energy production, and reduced operational costs.

¹⁴⁸ Nitsche, C., Schroedl, S., & Weiss, W. (2004, June). Onboard diagnostics concept for fuel cell vehicles using adaptive modelling. In *IEEE Intelligent Vehicles Symposium, 2004* (pp. 127-132). IEEE. DOI: [10.1109/IVS.2004.1336368](https://doi.org/10.1109/IVS.2004.1336368)

¹⁴⁹ Subotić, V., & Hochenauer, C. (2022). Analysis of solid oxide fuel and electrolysis cells operated in a real-system environment: State-of-the-health diagnostic, failure modes, degradation mitigation and performance regeneration. *Progress in Energy and Combustion Science*, 93, 101011. <https://doi.org/10.1016/j.pecs.2022.101011>.

¹⁵⁰ Ghorbani, M., Boley, M., Nakashima, P. N. H., & Birbilis, N. (2024). An active machine learning approach for optimal design of magnesium alloys using Bayesian optimisation. *Scientific Reports*, 14(1), 8299. <https://doi.org/10.1038/s41598-024-59100-9>.

¹⁵¹ Zhang, X., & Pisu, P. (2014). An unscented Kalman filter based on-line diagnostic approach for PEM fuel cell flooding. *International Journal of Prognostics and Health Management*, 5(1). <https://doi.org/10.36001/ijphm.2014.v5i1.2204>.

¹⁵² https://ocf.co.uk/media/ogjnbj3h/33315_oilandgasexplainer_eguide.pdf
<https://www.redoakconsulting.co.uk/energy-utilities/>
<https://aspioneer.com/why-high-performance-computing-is-vital-for-oil-gas-exploration/>

- **Predictive Maintenance:**
By processing data from intelligent sensors on energy assets (e.g., turbines, pipelines), HPC supports predictive maintenance, reducing downtime and extending equipment life.
- **Smart Grid Management:**
HPC powers real-time data analysis for grid operations, demand forecasting, and outage response, enhancing grid resilience and efficiency.
- **Renewable Energy Integration:**
HPC analyzes weather patterns and historical data to optimize the integration of wind, solar, and tidal energy into the grid, improving resource allocation and grid stability.
- **Cybersecurity:**
Advanced analytics running on HPC platforms detect and respond to cyber threats in real time, protecting critical infrastructure.

Industry Examples

Company/ Project	Application Area	HPC Role/Benefit
Siemens Gamesa	Wind Farm Optimization	Digital twins for layout optimization, boosting output and reducing costs
QED Naval	Tidal Turbine Deployment	CFD simulations for cost-effective, efficient deployment
Oil & Gas Majors	Reservoir Simulation	Seismic data processing, extraction optimization, risk reduction
Utilities	Smart Grid & Demand Forecasting	Real-time analytics for grid management and demand response

Benefits of HPC in Energy

- Accelerates R&D and innovation cycles
- Reduces operational costs and energy waste
- Improves safety and environmental performance
- Enables data-driven decision-making at scale
- Supports the transition to cleaner, more sustainable energy systems

Challenges and Considerations

- **Data Volume and Complexity:** Managing and processing exabytes of data requires robust, scalable HPC infrastructure.
- **Platform Complexity:** Integrating and maintaining HPC systems across decentralized operations can be challenging.
- **Continuous Upgrades:** Rapid advances in algorithms and hardware necessitate ongoing investment in infrastructure.
- **Energy Efficiency:** As computational demands grow, so does the need for energy-efficient HPC solutions.

HPC is indispensable for the modern energy sector, driving efficiency, sustainability, and innovation. Its ability to handle massive data volumes and complex simulations underpins advancements in renewables, oil & gas, and grid management, positioning HPC as a critical enabler of the global energy transition. In the rapidly evolving energy sector, the intersection of High-Performance Computing

(HPC), Big Data, and Artificial Intelligence is unlocking unprecedented opportunities. As the demand for cleaner, more efficient, and intelligent energy systems grows, so does the complexity of the data and computations required to power innovation. AI then steps in to make sense of it all. Whether it's predicting equipment failure, optimizing renewable energy integration, or designing smart energy systems, AI algorithms—trained on HPC-powered simulations and big datasets—are reshaping how we produce, store, and consume energy.

2.3.1.2 HPC for AI Overview

It is difficult to design an HPC for AI loads coupled with simulation software, because:

- Two infrastructure perspectives: HPC (scheduler, digital twins, development intensive) and cloud (Kubernetes, AI-loads, deployment intensive).
- A volatile AI software stack, also a mix of proprietary and open source, especially at the workflow level.

However, typically, an HPC-AI platform is designed according to the following.

HPC-AI aims to support two partitions. The main partition (Training) will support both HPC-like and Kubernetes access methods. Another partition, much smaller, supports higher-level application access, to deploy AI-as-a-service applications. Both need to be multi-tenant (with appropriate security measures). The Training partition should also support user-managed application deployment software.

2.3.1.3 Infrastructure aspects

Compute:

Most computation power will come from accelerators (GPUs).

The nodes are split across two partitions: Training partition, for batch-oriented loads and a smaller AI-as-a-service partition, for online loads.

The hardware is identical in both partitions: 4 GPUs per node, 192MB vRAM for each GPU (HBM3e or better), precision BF16/FP16 (16 bit), FP8 (8 bit) and FP4 (4 bit).

Storage:

The AI-as-a-service partition: NVMe drives are exposed as Object storage. These blocks can be accessed via S3, and POSIX.

The Training partition: relies mainly on a parallel shared filesystem (NVMe based) 1TB/s for write, x10 for reads. The nodes in this partition will also have access to the Object storage (much faster metadata access, needed for many files).

Regarding encryption, the following is usually foreseen:

- Encryption at rest with Self-Encrypting Drives (SED)
- Encryption per-volume/per-tenant
- Encryption with end-user provided keys at least for object data storage pools.

Finally, each GPU-accelerated node includes local NVMe storage for fast, temporary, data access.

Networking:

External networking: 2 x 100 Gbps¹⁵³, which corresponds to the current requirements for HPC access¹⁵⁴. Later, a connection to EuroHPC Hyper-connectivity [4] service at 2 x 400 Gbps, as the HPC access needs are projected to reach around 350 Gbps by 2025.

Interconnect (intra-HPC networking) typically relies on a high-performance interconnect fabric with 4 links of 400Gb/s each, ensuring very high bandwidth (1.6Tb/s per server, or better). The fabric should have RDMA capabilities, which may be provided through RoCE.

2.3.1.4 *Software aspects*

- software tools for AI/ML model and application hosting at scale,
- extensive platforms for AI/ML end-to-end lifecycle management,
- CI/CD pipelines,
- data management tools including preparation and ETL,
- ML workflow management tools,
- containerized AI microservices designed for high-performance environments.

2.3.1.5 *Access interface*

Application access can either use a preconfigured access method or opt to reserve nodes and deploy its own access method.

Preconfigured access methods include:

- WebUI-based tools for interactive development, data exploration and computing, such as JupyterLab and Open OnDemand.
- AI-as-a-service partition: MLOps interface (KubeFlow, MLFlow), as a SaaS sandbox for end users.
- Training partition: Slurm and Kubernetes access.

In addition to these foundation access layers, several nodes will be dedicated to running specialized services, such as Slurm REST API server, Kubernetes deployment service, and any other required server to provide application-level access.

As part of this project, a more user-friendly interface will be preferred. The specific context of the EnergyGuard load allows for a specific, simpler interface.

2.3.2 **AI-HPC Convergence**

The convergence of Artificial Intelligence (AI) and High-Performance Computing (HPC) represents a transformative shift in computational science, enabling breakthroughs across diverse fields. This synergy leverages the computational power of HPC to enhance AI's capabilities, while AI optimizes HPC workflows, creating a mutually beneficial relationship. State-of-the-Art encompasses the following topics (which will be investigated in Task 3.5):

Heterogeneous Architectures: SoA HPC systems integrate GPUs, TPUs, FPGA and very soon emerging Quantum Processing Units (QPUs) to accelerate AI workloads (in conjunction with

¹⁵³ Meluxina-AI invitation to tender <https://ted.europa.eu/en/notice/-/detail/239719-2025>

¹⁵⁴ EuroHyPerCon <https://op.europa.eu/en/publication-detail/-/publication/122922c3-fd63-11ef-b7db-01aa75ed71a1/language-en>

software acceleration techniques). These architectures support deep learning models and complex simulations, pushing the boundaries of computational efficiency.

Scalable AI Models: Advances in distributed training techniques allow AI models to scale across thousands of nodes in HPC systems. This scalability is crucial for processing massive datasets in fields like genomics, climate modelling, and astrophysics. Distributed learning also appears in this topic as an important technique for optimal acceleration of the learning process in an HPC environment.

AI-Driven HPC Optimization: AI algorithms are increasingly used to optimize HPC operations, such as resource allocation, energy efficiency, and fault prediction. This integration reduces operational costs and enhances system reliability.

Despite these advancements, challenges remain. Energy consumption, data privacy, and the complexity of integrating AI with HPC systems are critical issues. Future research aims to develop energy-efficient algorithms, enhance data security, and simplify while optimizing the deployment of AI-HPC solutions.

2.3.3 Trusted and sovereign data sharing

As EnergyGuard moves toward establishing a robust infrastructure for AI-powered energy innovation, adopting a data space architecture provides the necessary foundation for secure, scalable, and interoperable data exchange. The following key benefits highlight why data space integration is essential to the success of the EnergyGuard initiative. These criteria also form the basis for evaluating the connectors summarized in Table 1.

- **Security & Data Sovereignty:** Data spaces ensure secure and controlled data exchange between stakeholders while complying with GDPR and data sovereignty principles. This is critical for protecting sensitive energy data, including grid stability information and consumer usage patterns.
- **Interoperability:** By facilitating cross-sector integration, data spaces enable communication between energy systems and other platforms, such as IoT devices and smart home infrastructures. Standards like NGSI-LD and GAIA-X promote smooth, protocol-based interactions between heterogeneous systems
- **Data Transparency & Trust:** With features like data provenance, auditing, and smart contracts, data spaces provide mechanisms for transparent and verifiable data exchanges, ensuring trust among stakeholders.
- **Scalability & Flexibility:** Data spaces offer scalable, modular architectures. Cloud-to-edge federation capabilities allow flexible data processing and dynamic adaptation as the EnergyGuard ecosystem expands.
- **Innovation & Collaboration:** These infrastructures foster innovation by enabling multi-stakeholder collaboration and integration of emerging technologies such as AI and blockchain, essential for smart grid and advanced energy applications.
- **Real-Time Analytics:** Real-time data access supports predictive analytics, demand response, and intelligent grid management. AI/ML models can operate effectively on the timely data flows enabled by data spaces.
- **Cost Efficiency:** Data spaces reduce development and integration costs by offering shared frameworks for data exchange. Improved energy data handling also supports demand optimization, reducing operational costs for stakeholders.

Feature/Benefit	Eclipse Dataspace Connector (EDC)	FIWARE Data Space Connector	TNO Security Gateway (TSG) Connector *	TRUE Connector	SOVITY Open-Source EDC Connector	Trusted Connector by Fraunhofer
Maintainer	Eclipse Foundation	FIWARE Foundation	TNO	Engineering Ingegneria Informatica SpA	Sovity GmbH	Fraunhofer AISEC
GitHub Project	https://github.com/eclipse-edc/Connector	https://github.com/FIWARE/data-space-connector	https://gitlab.com/tno-tsg/archive/helm-charts/connector	https://github.com/Engineering-Research-and-Development/true-connector	https://github.com/sovity/edc-ce	https://github.com/Fraunhofer-AISEC/trusted-connector
License	Open Source (Apache 2.0)	Open Source (MIT)	Open Source	Open Source (AGPLv3)	Open Source (Apache 2.0)	Open Source (Apache 2.0)
Maturity	TRL 8-9	TRL 6-7	TRL 8	TRL 6 (Sandbox)	TRL 9	TRL 7
Deployment Status	Cloud, On-Premises, Hybrid	Cloud, On-Premises	Cloud	Edge, Cloud, On-Premises, IoT	Cloud, On-Premises	Cloud, Edge, On-Premises
Service Level	Self-Service, Framework	Connector as a Service, Self-Service	Connector as a Service	Platform as a Service	Connector as a Service	Connector as a Service
Communication Protocol	Dataspace Protocol 2024-01 (HTTPS)	HTTPS (NGSI-LD)	IDS Multipart	IDS Multipart, IDSCP	Dataspace Protocol 0.8, 1.0 (HTTPS)	Not specified
GUI	No	Yes	Yes	Yes	Yes	Yes
Identity Management	No	Yes (Decentralized, DID/VCs)	No	No	Yes (Centralized, Decentralized)	Yes (Centralized)

Integration with Clearing House	No	Yes (Gaia-X Digital Clearing House)	No	No	No	No
Information model	No	Yes (IDS 4.1.0)	No	No	No	Yes (IDS 4.1.0)
Vocabulary	No	No	No	No	Yes (DCAT, ODRL, DSpace)	No
Compliance with EU Data Standards	GAIA-X, IDSA	FIWARE/NGSI-LD	IDSA	Energy Regulations, GDPR, Part of the FIWARE ecosystem, Uses the IDS Information Model	GAIA-X, IDSA	IDSA
Usage Control (Policy)	No	Yes (Context Broker)	Yes (Data Consumer, Security Level, Time Interval, Location, Number of usages)	Yes (IDS Connector, Security Level, User Role, Time Interval, Duration, Purpose, Number of usages)	Yes (Connector Restriction, Time Interval)	Yes (Application inside a Connector » Time Interval)
Usage in Projects	SmartAgriHubs, Gaia-X, Catena-X, Eona-X, Health-X DataLOFT	Barcelona Smart City, FIWARE Community, SCC, DOME Marketplace, i4Trust	SCSN, EU Blockchain Observatory, Enershare	MARKET4.0, AI REGIO, EUR3KA, PLATOON, MUSKETEERS, ONENET, Enershare	Catena-X, Omega-X, Mobility Data Space	-
Potential Drawbacks	Complex integration and setup, due to high customizability	- Reliant on FIWARE ecosystem - Complex setup for non-FIWARE systems	- Niche focus on security, not on data modeling, ontologies,	- Limited ecosystem and adoption - Focus on compliance over scalability	- Limited ecosystem and adoption - Focus on compliance over scalability	No built-in integration with energy domain standards.

	- Requires a deeper understanding of GAIA-X and IDS standards.	- Resource-heavy - Focused on smart cities and mobility—may require extra adaptation for energy sector needs.	or advanced analytics - Not optimized for high-volume data sharing - Less flexible for cross-sectoral integration	- Limited protocol support - May not be ideal for IoT and real-time data integration	- Surprisingly bad state of documentation - Proprietary elements may reduce flexibility for open research collaboration	May require customization.
--	--	--	---	---	--	----------------------------

Table 1: Data Space Connectors Overview

The current landscape of dataspace connectors offers several strong candidates, each with distinct strengths depending on project needs. The **Eclipse Dataspace Connector (EDC)** stands out for its broad ecosystem and modular architecture, making it ideal for projects that require flexibility and customization, though it may pose integration complexity. **FIWARE**, built around the NGSI-LD protocol, is particularly effective in IoT and smart city contexts but can be challenging to adopt outside its ecosystem due to infrastructure demands. The **TNO Security Gateway (TSG) Connector** emphasizes security and regulatory compliance, especially for industrial and blockchain-related use cases, but its specialized focus may limit adaptability for broader interoperability or real-time data needs. The **TRUE Connector** offers strong compliance and privacy features, but its scalability and adoption in large-scale energy scenarios remain limited, which could affect long-term sustainability and support. In contrast, the **Sovity Open-Source EDC Connector** emerges as a highly mature option (TRL 9), combining advanced policy enforcement, identity management, and compatibility with modern dataspace protocols—positioning it as a robust and secure solution for cross-sectoral data integration. The EnergyGuard consortium will also leverage the ENERSHARE data space to support secure and policy-aligned data exchange between stakeholders. Using available dataspace connector technologies, the integration aims to enable controlled sharing of energy-related datasets while respecting data sovereignty and compliance requirements. This approach will rely on existing infrastructure and tools from EU initiatives, with a focus on enabling interoperability rather than developing entirely new frameworks.

2.3.4 Big Data management

The state-of-the-art in the exploitation of Big Data using SQL-based tools revolves mainly around advancements in scalability, performance optimization, and integration with modern data ecosystems. Here are some key aspects:

- Distributed Query Execution: Many tools like Presto excels in executing SQL queries across distributed data sources, enabling real-time analytics on massive datasets. Their ability to query data from Hadoop, S3, and other storage systems without requiring data movement is a significant advantage.
- Performance Enhancements: Recent developments focus on optimizing query execution through techniques like dynamic filtering, adaptive query planning, and columnar storage formats. These improvements reduce significantly latency and enhance throughput.

- **Integration with Data Lakes:** The compatibility with modern data lake architectures allows seamless querying of structured and semi-structured data. This integration supports diverse use cases, from business intelligence to machine learning.
- **Federated Querying:** The ability to query multiple data sources simultaneously, including relational databases and NoSQL systems, is a cutting-edge feature. This federated approach simplifies data analysis across heterogeneous environments
- **Scalability and Elasticity:** Advances in cluster management and resource allocation enable tools like Presto to scale dynamically based on workload demands, ensuring efficient utilization of computational resources.

These developments position Presto and similar SQL-based tools as main direction R&D directions of WP3/Task 3.3 in harnessing the power of Big Data for actionable insights.

2.4 Cybersecurity and trustworthy AI Assessment

2.4.1 Adversarial attacks and AI defence mechanisms

As artificial intelligence becomes increasingly integrated into real-world applications—from autonomous vehicles and medical diagnostics to finance and cybersecurity—the importance of ensuring its robustness against malicious manipulation has never been greater. One of the most pressing concerns in this context is the phenomenon of adversarial attacks: subtle, often imperceptible changes made to input data that cause machine learning models to make incorrect predictions.

These attacks exploit vulnerabilities in AI models in surprisingly effective ways. For example, a slightly altered image that appears identical to the human eye might cause a neural network to misclassify a stop sign as a speed limit sign. Such vulnerabilities aren't just theoretical—they have been demonstrated in both controlled environments and physical-world scenarios, raising alarms about AI safety across industries¹⁵⁵.

Recent literature has identified various types of adversarial attacks. Evasion attacks occur during model inference—where inputs are subtly modified to “fool” the system—while poisoning attacks manipulate the training data itself, compromising the model from the ground up. Others, like model inversion and extraction attacks, aim to reconstruct sensitive training data or steal model parameters¹⁵⁶.

These attacks share a dangerous feature: transferability. That is, adversarial examples crafted for one model often work against entirely different models as well, even when those models differ in architecture. This property makes black box attacks particularly concerning, since attackers don't need direct access to the target model¹⁵⁷.

To counter these threats, the AI community has developed a growing set of defence mechanisms. One of the most established is adversarial training, which involves retraining models using adversarial examples to improve their resilience. While often effective, this approach can be computationally

¹⁵⁵ Adversarial Attacks and Defenses in Deep Learning" - Kiu Ren, Thianhang Zheng, Zan Quin, Xue Liu, 2020, <https://www.sciencedirect.com/science/article/pii/S209580991930503X?via%3Dihub>

¹⁵⁶ A Comprehensive Review of Adversarial Attacks in Machine Learning - Abdirashid Abukar Ahmed, Dr. Nirvair Neeru, 2024, <https://ijsrem.com/download/a-comprehensive-review-of-adversarial-attacks-in-machine-learning/>

¹⁵⁷ "Adversarial Attacks and Defenses in Deep Learning" - Kiu Ren, Thianhang Zheng, Zan Quin, Xue Liu, 2020, <https://www.sciencedirect.com/science/article/pii/S209580991930503X?via%3Dihub>

intensive. Other strategies involve preprocessing inputs to strip away potential perturbations, or masking gradients to make attacks harder to optimize. However, some of these approaches offer only temporary protection and can be circumvented by more sophisticated adversaries¹⁵⁸.

More innovative approaches are beginning to emerge. For instance, some researchers have proposed defence strategies inspired by cryptographic principles, treating the defence mechanism as a "secret key" that attackers cannot reverse-engineer¹⁵⁹. Others advocate for stateful detection, where the system monitors a sequence of user interactions to detect when someone is systematically probing for weaknesses, rather than relying on static, per-input defences¹⁶⁰.

The research landscape is also beginning to embrace interdisciplinary strategies. Some work focuses on causal inference techniques to better understand why models make errors under attack and how to trace these failures back to their root causes¹⁶¹. Others emphasize the role of explainable AI in uncovering the inner workings of models and shedding light on their vulnerabilities¹⁶².

Despite these advances, defending against adversarial attacks remains a cat-and-mouse game. As defences become more sophisticated, so too do the attacks. The path forward likely involves a combination of techniques—ranging from adversarial training and cryptographic protections to regulatory frameworks and model interpretability—to ensure AI systems are not only powerful, but also resilient and trustworthy¹⁶³.

2.4.2 Bias, Fairness, and Explainability in AI for Energy

AI and ML have become essential tools for handling large datasets in energy consumption analysis, allowing both the optimization of efficiency and the identification of disparities in access to and distribution of energy infrastructure¹⁶⁴. AI technologies can be used to efficiently improve the

¹⁵⁸ "Adversarial Attacks and Defenses in Machine Learning-Powered Networks: A Contemporary Survey" - Yulong Wang, Tong Sun, Shenghong Li, Xin Yuan, Wei Ni, Ekram Hossain, H. Vincent Poor, 2023, <https://arxiv.org/abs/2303.06302>

¹⁵⁹ "Bridging Machine Learning and Cryptography in Defence Against Adversarial Attacks" - Olga Taran, Shideh Rezaeifar & Slava VoloShinovskiy, 2019, https://link.springer.com/chapter/10.1007/978-3-030-11012-3_23

¹⁶⁰ "Stateful Detection of Black-Box Adversarial Attacks" - Steven Chen, Nicholas Carlini, D. Wagner, 2020, <https://dl.acm.org/doi/10.1145/3385003.3410925>

¹⁶¹ "Causal Effects of Adversarial Attacks on AI Models in 6G Consumer Electronics" - Da Guo, Zhengjie Feng, Zhen Zhang, Fazlullah Khan, Chien-Ming Chen, Ruibin Bai, 2024, <https://ieeexplore.ieee.org/document/10648590/authors#authors>

¹⁶² "Study on Adversarial Attacks Techniques, Learning Methods and Countermeasures: Application to Anomaly Detection" - Anis Bouaziz, Manh-Dung Nguyen, Valeria Valdés, Ana Cavalli, Wissam Mallouli, 2023, <https://www.scitepress.org/Link.aspx?doi=10.5220/0012125100003538>

¹⁶³ "Trustworthy AI Principles to Face Adversarial Machine Learning: A Novel Study" - CH.E.N. Sai Priya , Manas Kumar Yogi, 2023, <https://irojournals.com/aicn/article/view/5/3/2>

¹⁶⁴ Chien-fei Chen, Rebecca Napolitano, Yuqing Hu, Bandana Kar, Bing Yao, Addressing machine learning bias to foster energy justice, Energy Research & Social Science, Volume 116, 2024, 103653, ISSN 2214-6296, <https://doi.org/10.1016/j.erss.2024.103653>

(<https://www.sciencedirect.com/science/article/pii/S2214629624002445>)

effectiveness of energy systems monitoring, controlling and optimizing¹⁶⁵. In the management of energy and utilities, AI systems have been used to forecast energy consumption across electric grids, decreasing needless energy generation¹⁶⁶, estimate households' energy usage to provide information on personalized tariffs¹⁶⁷ and forecast maintenance requirements and faults within energy generation and distribution systems to provide information on proactive action¹⁶⁸.

In general, bias is considered a potential harm caused by AI systems because due to data driven technologies feature to acquire insights from the "*existing structures and dynamics of the societies they analyze, they can reproduce, reinforce and amplify the patterns of marginalization, inequality and discrimination that exists in these societies*"¹⁶⁹. Biases in ML-based energy systems are likely to cause among others inequality and discrimination, resulting in unequal resources access and unfair treatment of some communities, prolonging the effects of an initial factor, amplifying the original issue and resulting in unfair access to advanced technology and exclusion from economic opportunities. As analyzed by Chien-fei Chen et al, they can derive from various sources, starting from the data collection [in data collection (i.e. biases from raw data acquisition), integration (and preprocessing)], to algorithmic development (design of a model, feature selection and adopted evaluation metrics) and deployment process (data distribution of the training model is different from its actual deployment environment¹⁷⁰). More specifically, during data collection the following types of data bias have been identified, namely; sampling (failure to precisely represent the characteristics of the sum of the population during acquisition¹⁷¹ or during preprocessing when addressing missing data), measurement (when analysis is distorted from often inaccurate data), selection (biases originating from the criteria of data subject selected for training data which relate to particular characteristics of community), aggregation (when data unified at a higher level lead to unintentional presumptions for subjects of the

¹⁶⁵ Md Morshed Alam et al, Artificial intelligence integrated grid systems: Technologies, potential frameworks, challenges and research directions, renewable and sustainable Energy Reviews 211 (2025) 115251, <https://doi.org/10.1016/j.rser.2024.115251>

¹⁶⁶ Daut, M. A. M., Hassan, M. Y., Abdullah, H., Rahman, H. A., Abdullah, M. P., & Hussin, F. (2017). Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review. Renewable and Sustainable Energy Reviews, 70, 1108-1118. <https://doi.org/10.1016/j.rser.2016.12.015>

¹⁶⁷ Leslie, D., Rincón, C., Briggs, M., Perini, A., Jayadeva, S., Borda, A., Bennett, S.J. Burr, C., Aitken, M., Katell, M., Fischer, C. (2023). AI Ethics and Governance in Practice: An Introduction. The Alan Turing Institute.

¹⁶⁸ Leslie, D., Rincón, C., Briggs, M., Perini, A., Jayadeva, S., Borda, A., Bennett, S.J. Burr, C., Aitken, M., Katell, M., Fischer, C. (2023). AI Ethics and Governance in

Practice: An Introduction. The Alan Turing Institute.

¹⁶⁹ Leslie, D. (2019). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute. <https://doi.org/10.5281/zenodo.3240529>

¹⁷⁰ Chien-fei Chen, Rebecca Napolitano, Yuqing Hu, Bandana Kar, Bing Yao, Addressing machine learning bias to foster energy justice, Energy Research & Social Science, Volume 116, 2024, 103653, ISSN 2214-6296, <https://doi.org/10.1016/j.erss.2024.103653>. (<https://www.sciencedirect.com/science/article/pii/S2214629624002445>)

¹⁷¹ Big-data for building energy performance: Lessons from assembling a very large national database of building energy use Paul A. Mathew et al, Applied energy 140 (2015) 85-93

group¹⁷²) and representation in terms of region and time¹⁷³. An example of bias in model development regarding evaluation metrics is when chosen metrics are not consistent with the development of the model. Bias during model deployment can be split into domain shift and explanation bias such as faulty forecasts as a result of discrepancies in data subjects' attributes in training data and deployment phase and not enough transparency regarding the models' forecasts¹⁷⁴.

Explainability of an AI system is *"the level to which the processes and the rationale behind outcomes of the system can be understood by human users. This can include the extent to which the inner workings of the model can be transformed into plain language, to promote better decision making and trust"*¹⁷⁵. As mentioned by Leslie, various machine learning models produce their results by operating on high dimensional correlations that are beyond the human interpretative capabilities of reasoning. In these cases, the reasoning with regards to algorithmically produced results that impact decision subjects remains unclear to those individuals. Although in some use cases, this lack of explainability may be tolerable, in some applications, where the processed data may contain elements of discrimination, bias, inequity, or unfairness, the lack of transparency of the model may cause great challenges¹⁷⁶. As mentioned by Md Morshed Alam et al in the article "Artificial intelligence integrated grid systems: Technologies, potential frameworks, challenges, and research directions", intelligence systems that utilize AI algorithms such as DL, ML, and DRL are often considered "black boxes" due to their complicated functionality and process of decision-making. As a result, it is difficult to trust these systems to ensure that they are making decisions in a user-friendly way¹⁷⁷.

*"Fairness can be defined in various ways, but it can be expressed by the extent to which an AI system promotes or prevents the mitigation of bias and the exclusion of discriminatory influences on its outputs and implementations"*¹⁷⁸. Biased machine learning systems may reinforce discrimination based on race/ethnicity, gender and age, thus worsening inequalities in society. An indicative example is systems

¹⁷² Evaluating potential sources of aggregation Bias with a structural optimization model of the U.S. Forest Sector, C.M. Eade et al J. For Econ (2019)

¹⁷³ Chien-fei Chen, Rebecca Napolitano, Yuqing Hu, Bandana Kar, Bing Yao, Addressing machine learning bias to foster energy justice, Energy Research & Social Science, Volume 116, 2024, 103653, ISSN 2214-6296, <https://doi.org/10.1016/j.erss.2024.103653>.(<https://www.sciencedirect.com/science/article/pii/S2214629624002445>)

¹⁷⁴ Chien-fei Chen, Rebecca Napolitano, Yuqing Hu, Bandana Kar, Bing Yao, Addressing machine learning bias to foster energy justice, Energy Research & Social Science, Volume 116, 2024, 103653, ISSN 2214-6296, <https://doi.org/10.1016/j.erss.2024.103653>.(<https://www.sciencedirect.com/science/article/pii/S2214629624002445>)

¹⁷⁵ Leslie, D., Burr, C., Aitken, M., Cowls, J., Katell, M., and Briggs, M. (2021). Artificial intelligence, human rights, democracy, and the rule of law: a primer. The Council of Europe.

¹⁷⁶ Leslie, D. (2019). Understanding artificial intelligence ethics and : A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute. <https://doi.org/10.5281/zenodo.3240529>

¹⁷⁷ <https://doi.org/10.1016/j.rser.2024.115251>

¹⁷⁸ Leslie, D., Burr, C., Aitken, M., Cowls, J., Katell, M., and Briggs, M. (2021). Artificial intelligence, human rights, democracy, and the rule of law: a primer. The Council of Europe.

that unintentionally discriminate on the basis of race in determining credit eligibility for energy sufficiency loans or financing for clean energy that can result in disruptive injustice (meaning energy justice that is centered on the ethical principles that should rule resource distribution among the members of a society), impeding marginalized group fair access to these resources¹⁷⁹.

2.4.2.1 *Bias in Energy AI Systems*

Bias can arise in energy related AI and ML systems in various stages of the ML lifecycle – from data collection to model deployment¹⁸⁰, as explained in the following sections.

Bias in data collection

Firstly, bias can be related to data collection. Sampling bias is one such case, which occurs when the samples of the dataset chosen do not accurately reflect the characteristics of the entire data / population. For example, Haben et. al.¹⁸¹ show that 42% of all reproducible Low Voltage (LV) load forecasting papers found use the same Irish CER smart-meter trial; the houses in that trial are mostly 3- or 4-bedroom homes, so methods calibrated there may not be representative of the wider energy system. Over-reliance on that convenient dataset skews model development and produces algorithms that transfer poorly to other feeder types or climates.

Another potential bias source is measurement bias, which results in data being inaccurate or incomplete due to limited access to data or missing / faulty sensors. Ming et. al.¹⁸² illustrate one way to avoid this bias in smart meter data by detecting the inaccurate sensors using historical data and an LSTM and CNN based method and replacing them.

Aggregation bias occurs when aggregating data of individuals (for example by averaging) and using this data to make unintentional assumptions about the whole group. An example of this kind of problem is the Modifiable Areal Unit Problem (MAUP), which occurs when data is aggregated using different geographic boundaries, resulting in a substantial shift in the resulting patterns and metrics¹⁸³. Power-outage records, for example, are usually reported by utilities according to their own service-area boundaries, which rarely align with county lines and often cut across multiple ZIP codes. To use these records for restoration planning or resource allocation at the county or census-tract level, analysts have to re-aggregate or disaggregate the data—steps that can unintentionally sideline certain communities and intensify energy-justice concerns¹⁸⁴.

Selection bias arises from how subjects are chosen or included in a study. For example, in U.S. national survey data such as the Residential Energy Consumption Survey (RECS)¹⁸⁵, deliberately omitting vacant homes and certain types of housing units can lead to selection bias.

¹⁷⁹ Chien-fei Chen, Rebecca Napolitano, Yuqing Hu, Bandana Kar, Bing Yao, Addressing machine learning bias to foster energy justice, Energy Research & Social Science, Volume 116, 2024, 103653, ISSN 2214-6296,

<https://doi.org/10.1016/j.erss.2024.103653>(<https://www.sciencedirect.com/science/article/pii/S2214629624002445>)

¹⁸⁰ <https://doi.org/10.1016/j.erss.2024.103653>

¹⁸¹ <https://doi.org/10.1016/j.apenergy.2021.117798>

¹⁸² 10.1109/MIE.2020.3026197

¹⁸³ <https://doi.org/10.1068/a231025>

¹⁸⁴ <https://doi.org/10.1016/j.esr.2021.100752>

¹⁸⁵ <https://www.eia.gov/consumption/residential/>

Representation bias can occur when energy data that was collected about a specific place or time is used to build a model to predict a different one. For example, Quesada et. al.¹⁸⁶ point out that most open smart meter datasets come from a handful of English-speaking countries, skewing ML models toward those contexts. Their Spanish dataset is presented to expand geographic diversity and reduce temporal biases of such datasets.

Bias in development and deployment

Bias can also be introduced during the model development and deployment stage of the ML lifecycle.

Algorithmic bias, also referred to as model bias, arises because of choices related to model development, such as architecture and feature selection¹⁸⁷. For example, Kalfountzou et. al.¹⁸⁸ report that dropping just one predictor—household income—cuts accuracy for energy-poor homes to 32%, while overall accuracy stays above 83%, illustrating how an apparently highly accurate model can become heavily biased when a key feature is omitted.

Bias can also emerge during the deployment phase of a model. Domain shift and explanation bias are two common types of biases that may arise at this stage¹⁸⁹. Domain shift occurs when the data the model was trained on differs from the data it is attempting to predict, while explanation bias is caused by the users of the models not understanding the way they arrived at their results. The latter type of bias can be tackled with the use of explainable AI algorithms like NBEATS¹⁹⁰ or using more general-purpose explainability methods like SHAP¹⁹¹. EnergyGuard will consider utilising explainability methods as they can help detect all the types of bias, by interpreting the way input features affect the output of models. Tools that will be integrated into the platform such as DeepTSF¹⁹² have a working implementation of SHAP.

2.4.2.2 Fairness Considerations in Energy Systems

Fairness in AI for the energy sector focuses on ensuring that AI-driven decisions—such as pricing, grid optimization, or energy distribution—do not disproportionately disadvantage certain groups. While fairness and bias are closely related, fairness is a deliberate goal that requires context-aware design and attention to the diverse needs of stakeholders. Since it refers to the absence of bias, achieving fairness can be challenging due to the various types of bias that may arise in these systems. Despite growing recognition of the importance of fairness in the energy sector, evaluating how it is addressed in current studies remains difficult due to the lack of a uniform interpretation across the scientific literature. Nonetheless, fairness must be a critical consideration in the development and deployment of AI systems in the energy domain, as it involves creating systems that equitably distribute energy benefits, avoid discriminatory outcomes, and support just energy transitions, particularly for vulnerable and underserved communities. One strategy to ensure fairness is to address bias by improving the quality and diversity of the datasets used in model training. Energy data are usually imbalanced and

¹⁸⁶ <https://www.nature.com/articles/s41597-023-02846-0>

¹⁸⁷ <https://doi.org/10.1016/j.erss.2024.103653>

¹⁸⁸ <https://doi.org/10.3390/en18051133>

¹⁸⁹ <https://doi.org/10.1016/j.erss.2024.103653>

¹⁹⁰ [10.48550/arXiv.1905.10437](https://arxiv.org/abs/1905.10437)

¹⁹¹ <https://shap.readthedocs.io/en/latest/>

¹⁹² <https://doi.org/10.1016/j.softx.2024.101758>

underrepresented; thus, it is essential that all groups are adequately represented, with data collected from a wide range of sources to capture different contexts and experiences. Data augmentation techniques can also support this goal by generating synthetic data that enhances representation and helps reduce bias in AI system¹⁹³.

Another key strategy is developing fairness-aware algorithms. This approach focuses on creating AI models that explicitly consider fairness during their design. It can be achieved by integrating fairness constraints into the training process or by using algorithms specifically designed to reduce bias. For example, Luo et al.¹⁹⁴ explores how fairness and ethical principles can be embedded into smart grid optimization to promote equitable energy distribution. By integrating fairness constraints—such as proportional and equity-based fairness—and using Shapley values from game theory, the proposed optimization models ensure that energy is allocated more justly, especially for low-income and vulnerable populations. The findings highlight that these approaches can significantly reduce the energy poverty gap and support more inclusive access to energy.

Fairness has become a critical concern in demand response programs and energy pricing. AI algorithms used in these systems should be prevented from unintentionally favouring wealthier households with access to energy-efficient technologies, which could result in higher costs for lower-income households with limited participation opportunities. This highlights the need for fairness-aware algorithms that account for the socioeconomic differences between households. Likewise, biased AI in energy pricing may lead to higher charges for disadvantaged groups, worsening their financial strain. In this context, Thibodeau et al.¹⁹⁵ use reinforcement learning (RL) and bandits to help policy makers design fairer dynamic pricing systems that promote equitable demand across consumer groups. A key innovation, FairReplayBuffer, ensures the RL agent samples experiences uniformly across a discretized fairness space, allowing it to better address underrepresentation and improve social welfare outcomes.

An important energy-related field where fairness is crucial is energy resource allocation, especially with the rise of distributed energy resources like solar, wind, electric vehicles, and batteries. These allow users to generate, store, and sell electricity, often with the help of energy aggregators who negotiate on their behalf, distributing costs and benefits within the group. However, maximizing efficiency can lead to unequal resource distribution, favouring higher-income households or businesses. Li et al.¹⁹⁶ formalize the problem of fair energy allocation, providing a framework that explores the trade-offs between efficiency and fairness while optimizing resource distribution across individual users.

¹⁹³ Meiser, M.; Zinnikus, I. A Survey on the Use of Synthetic Data for Enhancing Key Aspects of Trustworthy AI in the Energy Domain: Challenges and Opportunities. *Energies* **2024**, *17*, 1992. <https://doi.org/10.3390/en17091992>

¹⁹⁴ Luo, D., & Pan, W. (2025). Ethical Considerations in Smart Grid Optimization: Promoting Energy Equity and Fairness. *International Theory and Practice in Humanities and Social Sciences*, *2*(2), 145–163. <https://doi.org/10.70693/itphss.v2i2.212>

¹⁹⁵ J. Thibodeau, H. Nekoei, A. Ta'ik, J. Rajendran, G. Farnadi, Fairness incentives in response to unfair dynamic pricing (2024). arXiv:2404.14620. <https://doi.org/10.48550/arXiv.2404.14620>

¹⁹⁶ J. Li, M. Motoki, B. Zhang, Balancing fairness and efficiency in energy resource allocations (2024). arXiv:2403.15616. <https://doi.org/10.48550/arXiv.2403.15616>

2.4.2.3 Explainability in AI for Energy

The energy sector is a critical infrastructure, with decisions significantly impacting the economy, environment and society. As a result, there is a growing need for comprehensive explanations of AI-driven decision-making processes. In this context, explainability analysis has emerged as a high-demand practice for understanding model predictions and mitigating risks. Among the most widely used techniques are SHAP and LIME, although a variety of XAI tools are employed to explain predictions across different energy-related fields. Energy forecasting is becoming an increasingly prominent area for applying explainability analysis. For instance, Lee et al.¹⁹⁷ utilized SHAP and Partial Dependence Plots (PDPs) to interpret load forecasting models. Similarly, Sakkas et al.¹⁹⁸ employed counterfactual explanations to make energy demand forecasting understandable. DeepTSF¹⁹⁹ is a comprehensive machine learning operations (MLOps) framework for time series forecasting that incorporates SHAP for explanation, offering insightful visualizations of feature contributions. Shajalal et al.²⁰⁰ proposed an explainable deep energy demand forecasting framework that combines LSTM-based models with DeepLIFT to approximate SHAP values. This enables transparent predictions for smart home users by highlighting the influence of temporal energy consumption attributes, such as appliance usage, area-specific consumption, activity types, and seasonal patterns.

In the domain of renewable energy forecasting, explainability is essential for gaining insights, managing the uncertainty of solar power generation, and developing intelligent demand response strategies. Sarp et al.²⁰¹ used the ELI5 tool to explain feature contributions in XGBoost-based PV energy forecasting models. In addition to ELI5, SHAP and LIME have also been utilized for PV power generation predictions, aiding in the elimination of irrelevant features²⁰². Further studies, such as Mitrentsis and Lens²⁰³, extend

¹⁹⁷ Y.-G. Lee, J.-Y. Oh, G. Kim, Interpretation of load forecasting using explainable artificial intelligence techniques, *The Transactions of the Korean Institute of Electrical Engineers* 69 (2020) 480–485.

doi:<https://doi.org/10.5370/kiee.2020.69.3.480>

¹⁹⁸ N. Sakkas, S. Yfanti, C. Daskalakis, E. Barbu, M. Domnich, Interpretable forecasting of energy demand in the residential sector, *Energies* 14 (20) (2021). doi:10.3390/en14206568.

¹⁹⁹ S. Pelekis, T. Pountridis, G. Kormpakis, G. Lampropoulos, E. Karakolis, S. Mouzakitis, D. Askounis, DeepTSF: Codeless machine learning operations for time series forecasting, *SoftwareX* 27 (2024) 101758.

doi:<https://doi.org/10.1016/j.softx.2024.101758>

²⁰⁰ M. Shajalal, A. Boden, G. Stevens, Forecastexplainer: Explainable household energy demand forecasting by approximating shapley values using deeplift, *Technological Forecasting and Social Change* 206 (2024) 123588.

doi:<https://doi.org/10.1016/j.techfore.2024.123588>.

²⁰¹ S. Sarp, M. Kuzlu, U. Cali, O. Elma, O. Guler, An interpretable solar photovoltaic power generation forecasting approach using an explainable artificial intelligence tool, in: *2021 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 2021, pp. 1–5. doi:10.1109/ISGT49243.2021.9372263

²⁰² M. Kuzlu, U. Cali, V. Sharma, G. Guler, Gaining insight into solar photovoltaic power generation forecasting utilizing explainable artificial intelligence tools, *IEEE Access* 8 (2020) 187814–187823.

doi:10.1109/ACCESS.2020.3031477

²⁰³ G. Mitrentsis, H. Lens, An interpretable probabilistic model for short-term solar power forecasting using natural gradient boosting, *Applied Energy* 309 (2022) 118473. doi:<https://doi.org/10.1016/j.apenergy.2021.118473>

explainability in PV generation forecasting. Solar irradiance prediction has also been approached using inherently interpretable models²⁰⁴, as well as SHAP, feature importance analysis, and other methods²⁰⁵. Recently, explainability analysis has seen increased integration into machine learning models focusing on buildings and energy efficiency, aiming to enhance transparency and generate actionable insights. These techniques have been applied to models predicting various outcomes such as energy consumption²⁰⁶, energy efficiency ratings²⁰⁷, heating and cooling loads²⁰⁸, and greenhouse gas (GHG) emissions²⁰⁹. These methods help identify the most influential features for each outcome, quantifying their impact and enabling the prioritization of retrofitting strategies based on the elements that most significantly affect building performance. They help identify the most influential features and quantify their impact, allowing stakeholders to prioritize retrofitting strategies based on the elements most affecting building performance. SHAP has been used to justify specific energy efficiency improvement recommendations²¹⁰, helping build trust and drive adoption of retrofit strategies. Additionally, similar explainability approaches have been employed to analyse the financial implications of building features and assess their economic benefits²¹¹.

²⁰⁴ H. Wang, R. Cai, B. Zhou, S. Aziz, B. Qin, N. Voropai, L. Gan, E. Barakhtenko, Solar irradiance forecasting based on direct explainable neural network, *Energy Conversion and Management* 226 (2020) 113487.

doi:<https://doi.org/10.1016/j.enconman.2020.113487>

²⁰⁵ M. Chaibi, E. M. Benghoulam, L. Tarik, M. Berrada, A. E. Hmaidi, An interpretable machine learning model for daily global solar radiation prediction, *Energies* 14 (21) (2021). doi:10.3390/en14217367

²⁰⁶ X. Cui, M. Lee, C. Koo, T. Hong, Energy consumption prediction and household feature analysis for different residential building types using machine learning and SHAP: Toward energy-efficient buildings, *Energy and Buildings* 309 (2024) 113997. doi:10.1016/J.ENBUILD.2024.113997.

X. Liu, H. Tang, Y. Ding, D. Yan, Investigating the performance of machine learning models combined with different feature selection methods to estimate the energy consumption of buildings, *Energy and Buildings* 273 (2022) 112408. doi:10.1016/J.ENBUILD.2022.112408.

²⁰⁷ M. Sun, C. Han, Q. Nie, J. Xu, F. Zhang, Q. Zhao, Understanding building energy efficiency with administrative and emerging urban big data by deep learning in Glasgow, *Energy and Buildings* 273 (2022) 112331. doi:10.1016/J.ENBUILD. 2022.112331.

²⁰⁸ B. S. Alotaibi, Advancing energy performance efficiency in residential buildings for sustainable design: Integrating machine learning and optimized explainable AI (AIX), *International Journal of Energy Research* 2024 (2024) 6130634. doi:10.1155/2024/6130634.

²⁰⁹ Y. Zhang, B. K. Teoh, M. Wu, J. Chen, L. Zhang, Data-driven estimation of building energy consumption and GHG emissions using explainable artificial intelligence, *Energy* 262 (2023) 125468. doi:10.1016/J.ENERGY.2022.125468

²¹⁰ P. Rempi, S. Pelekis, A. M. Tzortzis, E. Karakolis, C. Ntanos, D. Askounis, Explainable ai for building energy retrofitting under data scarcity (2025). arXiv:2504.06055. doi:[10.48550/arXiv.2504.06055](https://doi.org/10.48550/arXiv.2504.06055)

²¹¹ K. Konhäuser, T. Werner, Uncovering the financial impact of energy-efficient building characteristics with explainable artificial intelligence, *Applied Energy* 374 (2024) 123960. doi:10.1016/J.APENERGY.2024.123960

Explainability is also critical in energy system maintenance, where it supports fault detection, risk assessment, and proactive intervention²¹². A major application is in photovoltaic (PV) panel monitoring, where early-stage fault detection is critical to prevent severe power loss. Sairam et al²¹³ employed LIME to identify specific features, such as temperature anomalies or voltage drops, offering transparency in panel maintenance decisions. Dou et al²¹⁴ incorporated SHAP into reinforcement learning models used for renewable energy dispatch to explain how factors like energy availability influenced operational choices. This enables operators to proactively adjust strategies to ensure efficient energy distribution and avoid system overloads. Additionally, tree-based algorithms with SHAP have proven effective for fault diagnosis in grid-connected PV systems, offering interpretable insights into the nature of faults for more accurate detection and timely response²¹⁵.

2.4.3 Security and Governance Frameworks for AI

AI technologies continue to expand into sectors such as critical infrastructure, digital health, finance, and national security, the question of how to secure these systems is becoming more urgent than ever. AI models not only offer unprecedented capabilities but also introduce novel risks—ranging from algorithmic manipulation to misuse in autonomous decision-making. In response, researchers, governments, and industry leaders are increasingly focusing on the development of **comprehensive security frameworks** tailored specifically to the complexities of AI systems. These frameworks aim to move beyond conventional cybersecurity strategies by addressing the unique operational, ethical, and regulatory challenges posed by intelligent algorithms.

The evolution of security frameworks for AI has been shaped by the growing recognition of AI's unique vulnerabilities. Unlike traditional IT systems, AI models—particularly those based on machine learning—can be susceptible to adversarial attacks, data poisoning, model inversion, and misuse of autonomous decision-making capabilities. These threats demand specialized approaches that go beyond conventional cybersecurity tools.

Several efforts have been made to define structured approaches to securing AI systems. A notable example is the AI security framework proposed by Jing et al. (2021)²¹⁶, which organizes AI security around a graded capability model and emphasizes the integration of ethical principles and legal compliance into the technical development lifecycle. This type of layered framework ensures that

²¹² M. R. Shadi, H. Mirshekali, H. R. Shaker, Explainable artificial intelligence for energy systems maintenance: A review on concepts, current techniques, challenges, and prospects, *Renewable and Sustainable Energy Reviews* 216 (2025) 115668. doi:<https://doi.org/10.1016/j.rser.2025.115668>

²¹³ S. Sairam, S. Srinivasan, G. Marafioti, B. Subathra, G. Mathisen, K. Bekiroglu, Explainable incipient fault detection systems for photovoltaic panels (2020). arXiv:2011.09843. doi:<https://doi.org/10.48550/arXiv.2011.09843>

²¹⁴ J. Dou, X. Wang, Z. Liu, Z. Jiao, Y. Han, Q. Sun, J. He, Multi-objective explainable smart dispatch for integrated energy system based on an explainable mo-rl method, *Comput. Electr. Eng.* 118 (PB) (Sep. 2024). doi:10.1016/j.compeleceng.2024.109417

²¹⁵ H. N. Noura, Z. Allal, O. Salman, K. Chahine, Explainable artificial intelligence of tree-based algorithms for fault detection and diagnosis in grid-connected photovoltaic systems, *Engineering Applications of Artificial Intelligence* 139 (2025) 109503. doi:<https://doi.org/10.1016/j.engappai.2024.109503>.

²¹⁶ "An Artificial Intelligence Security Framework" - Huiyun Jing, Wei Wei, Chuan Zhou and Xin He, 2021, <https://iopsience.iop.org/article/10.1088/1742-6596/1948/1/012004>

security considerations are embedded from the ground up—starting from model design to deployment and post-deployment monitoring.

A parallel framework by Yu (2021)²¹⁷ introduces the idea of a security boundary model for AI, stressing the importance of defining clear perimeters around where and how AI systems operate, to manage both internal and external threats.

Security frameworks are also increasingly addressing governance dimensions: who controls the AI, how decisions are audited, and how responsibilities are distributed. The AI TRiSM (Trust, Risk, and Security Management) framework exemplifies this shift, providing a model that connects technical robustness, ethical transparency, and organizational accountability. It has seen application in sectors like healthcare and finance, where trust is paramount²¹⁸.

Similarly, Villegas-Ch. and Garcia-Ortiz (2023)²¹⁹ propose a comprehensive framework that combines data protection policies, technical safeguards, and regulatory compliance mechanisms. Their approach highlights how AI security must be viewed as a combination of legal, social, and technological systems rather than as isolated technical configurations.

While frameworks are maturing, several key challenges remain:

- **Dynamic threat landscape:** AI systems must adapt to evolving threats, including AI-driven cyberattacks, which themselves exploit machine learning models.
- **Interdisciplinary gaps:** Many frameworks lack clear integration between AI ethics, law, and technical implementation, making end-to-end accountability difficult to achieve.
- **Regulatory fragmentation:** The absence of unified global standards complicates international cooperation, particularly in cross-border applications like cloud AI and autonomous vehicles.
- **Trust and explainability:** Transparency is critical for public trust, yet many systems remain opaque ("black boxes"), making it hard to verify compliance or detect manipulation.

Researchers such as Jalili and Dziaikovskii (2023)²²⁰ also point to the need for frameworks that remain resilient under economic and geopolitical stress, integrating emerging tools like zero-knowledge proofs and federated governance models for state-level applications.

²¹⁷ "Security Framework of Artificial Intelligence System" - Rong Yu, 2021

<https://iopscience.iop.org/article/10.1088/1742-6596/1927/1/012011>

²¹⁸ "Artificial Intelligence Trust, Risk and Security Management (AI TRiSM): Frameworks, applications, challenges and future research directions" - Adib Habbal, Mohamed Khalif Ali, Mustafa Ali Abuzaraida, 2024,

<https://www.sciencedirect.com/science/article/abs/pii/S0957417423029445?via%3Dihub>

²¹⁹ "Toward a Comprehensive Framework for Ensuring Security and Privacy in Artificial Intelligence" - William Villegas, Joselin García-Ortiz, 2023

<https://www.mdpi.com/2079-9292/12/18/3786>

²²⁰ "State data security backed by Artificial Intelligence and Zero Knowledge Proofs in the context of sanctions and economic pressure" - Abdul Qawi Jalili, Anton Dziaikovskii, 2023

<https://ea21journal.world/index.php/ea.V202-01/>

2.5 AI Regulation and compliance

2.5.1 International Regulations and compliance frameworks

2.5.1.1 Background

A key issue in the AI debate is how to regulate it to ensure ethical and responsible use. In recent years, there has been a surge in AI regulation efforts at various levels of governance, with over 700 national initiatives recorded by the OECD AI Policy Observatory in 60 countries and regions. While early research focused on national efforts, since 2020 international organizations (IOs) have overtaken national authorities as the main promoters of regulatory initiatives, according to the Council of Europe²²¹.

Efforts to categorize the legal strategies for AI regulation have identified three main types, although their boundaries are not strictly defined²²²:

1. **Adaptation of existing legislation (overhauls):** This involves amending existing legal provisions to cover AI-related issues. For example, existing anti-discrimination laws may be updated to address discrimination resulting from the use of AI systems.
2. **Creation of new, specialized legislation:** These are entirely new regulations specifically tailored to AI. They vary in scale and scope and include laws focused on AI technologies, such as facial recognition or autonomous vehicles.
3. **Holistic (omnibus) legislation:** These are broader laws not limited to specific domains, often covering multiple substantive issues. Less common due to the time needed for design and adoption, these frameworks aim to regulate AI, rather than specific sectors or applications²²³.

The most comprehensive regulatory effort to date has been the European Commission's proposal for the Artificial Intelligence Act (AI Act), which aims to create a clear legal framework for the development and deployment of AI systems^{224,225}, finally published in July of 2024. It focuses on system usage by proposing a risk-based approach. The framework classifies AI applications by risk level, imposing stricter obligations on high-risk applications. It also outlines conformity assessment procedures prior to market entry and includes mechanisms for oversight and governance at both EU and national levels²²⁶.

Beyond the AI Act, the EU also has complementary legislation relevant to AI technologies, such as the General Data Protection Regulation (GDPR), which sets unified rules for personal data protection, and the Digital Services Act (DSA), which protects consumers and their fundamental rights online.

²²¹ The Global Governance of Artificial Intelligence: Next Steps for Empirical and Normative Research, International Studies Review, Tallberg et al, 2023

²²² A systematic review of regulatory strategies and transparency mandates in AI regulation in Europe, the United States, and Canada, Sloane, Mona and Wüllhorst, Elena, 2025

²²³ A systematic review of regulatory strategies and transparency mandates in AI regulation in Europe, the United States, and Canada, Sloane, Mona and Wüllhorst, Elena, 2025

²²⁴ Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation, Natalia Díaz-Rodríguez et al., Information Fusion (2023)

²²⁵ Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain Union legislative acts, Union, E, (2021)

²²⁶ Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation, Natalia Díaz-Rodríguez et al., Information Fusion (2023)

In contrast, China has rapidly developed AI technologies while using them to strengthen state surveillance, including facial recognition and social credit applications. Its policies center on using AI to bolster state control, with less emphasis on individual rights. China also seeks to shape international AI rules, though it faces obstacles due to its differing regulatory philosophy from dominant global models²²⁷.

The United States takes a more decentralized and sector-specific approach to AI regulation, resulting in a wide variety of guidelines created by states and federal agencies tailored to specific sectors. While this allows for flexibility and domain-specific legislation, it can also lead to regulatory gaps and inconsistencies. A key player is the National Institute of Standards and Technology (NIST), which plays a central role in developing standards and benchmarks for AI systems²²⁸.

It is therefore clear that AI regulation is a complex and evolving field, with varying approaches and strategies. As AI continues to advance and penetrate multiple areas of society and the economy, the need for effective, ethical, and collaborative regulation becomes increasingly critical. As mentioned, a growing body of principles, laws, policies, frameworks, declarations, voluntary commitments, standards and best practices is continually growing with regards to AI regulation and compliance.

In addition to that, documents providing guidance, issued by appropriate bodies and organizations contribute to clarification of requirements and complicated meanings used in the continuously evolving environment of AI. The documents can be divided according to their nature to those that are directly legally binding (hard law, meaning legally binding instruments laying down sanctions and being enforced through judicial proceedings²²⁹) such as regulations and directives (the latter subject to their national transposition) and those that are not legally binding such Guidelines and principles (soft law, non-legally binding, the infringement of which has with no legal sanctions²³⁰). In addition, it should be highlighted that AI principles such as OECD's AI Principles of UNESCO's Recommendation on Ethics of AI have the capacity to shape global standards, especially in those cases where national governments commit to voluntarily include such guidance into their national initiatives regarding AI governance ²³¹. In the European Union, an extensive AI governance framework is being under development. In addition

²²⁷ Human-Centric Versus State-Driven: A Comparative Analysis of the European Union's and China's Artificial Intelligence Governance Using Risk Management, International Journal of Intelligent Information Technologies, Arora et al, (2021)

²²⁸ Shaping the future of AI: balancing innovation and ethics in global regulation, Kashefi et al, Uniform Law Review (2024)

²²⁹ David Leslie, Christopher Burr, Mhairi Aitken, Josh Cowls, Mike Katell, & Morgan Briggs; With a foreword by Lord Tim Clement-Jones , Artificial intelligence, human rights, democracy, and the rule of law - A primer (2021), <https://edoc.coe.int/en/artificial-intelligence/10206-artificial-intelligence-human-rights-democracy-and-the-rule-of-law-a-primer.html>

²³⁰ David Leslie, Christopher Burr, Mhairi Aitken, Josh Cowls, Mike Katell, & Morgan Briggs; With a foreword by Lord Tim Clement-Jones , Artificial intelligence, human rights, democracy, and the rule of law - A primer (2021), <https://edoc.coe.int/en/artificial-intelligence/10206-artificial-intelligence-human-rights-democracy-and-the-rule-of-law-a-primer.html>

²³¹ IAPP, FTI Consulting technology, AI Governance in Practice Report 2024

to AI Act, in the energy sector, various regulations on data protection, cybersecurity and digital markets affect AI development and deployment²³².

The following analysis is divided in the analysis of the specific frameworks, regulations and guidelines related to AI and other legal relevant sectors when dealing with AI, regulation, standards and recommendation related to regulatory sandboxes and experimental framework and principles and guidance related to ethics.

2.5.1.2 *Specific frameworks, regulations and guidelines related to AI*

In this section documents that are related to the AI regulation and compliance are laid down and briefly analyzed, while at the end of this section, a table defining each documents category and indicating its relevance to the project is inserted for clarity reasons.

As a general remark, it should be underlined that the following regulatory analysis is primarily focused on EU legislation thus probable national legislations are not mentioned therein and when necessary, they will be examined and added in the next steps of the project. Moreover, it should be noted that regulation analysis below refers to AI legislation and compliance framework and does not extend to other compliance obligations or requirements of probable AI applications to be tested to the Energy Guard Platform in sectors other than AI. The list of regulations is not exclusive and sectoral legislation or guidelines may be relevant, if considered necessary, according to the requirements of the project (for example the EU Network Code in Cybersecurity for electricity designed for resilience in electricity grids²³³). Finally, this list is expected to be completed while the project proceeds, since regulations, implementing acts and guidelines probably related to the project are still expected, and the regulatory regime is still being set up. In addition, as mentioned in a study the analyzes the AI Act in the context of existing EU digital regulations and sectoral regulations, the AI Act is not complete yet since provisions of AI Act are to be defined and clarified more through so called technical standards drawn by expert committees²³⁴. Moreover, the General- Purpose AI Code of Practice, a voluntary tool aiming to assist providers of General-Purpose AI models to comply with the AI Act's specific obligations, was published on the 10th of July of 2025. The Code includes three Chapters, namely transparency and copyright for all providers of general-purpose AI models and chapter on Safety and Security for providers of general purpose AI models with systemic risk, is expected to be assessed for its adequacy by Commission and Member States²³⁵.

2.5.1.3 *Charter of Fundamental Rights of the European Union*

The EU Charter of Fundamental Rights (hereinafter "the Charter"²³⁶) became legally binding in December 2009 and includes the same values as the EU Treaties. The Charter constitutes primary EU law and as such it functions as a criterion for examining the validity of secondary EU legislation and national measures²³⁷. *"It brings together civil, political economic and social rights in a single text."*

²³² Jørgensen, B.N.; Ma, Z.G. Regulating AI in the Energy Sector: A Scoping Review of EU Laws, Challenges, and Global Perspectives. *Energies* 2025, 18, 2359. <https://doi.org/10.3390/en18092359>

²³³ https://energy.ec.europa.eu/news/new-network-code-cybersecurity-eu-electricity-sector-2024-03-11_en

²³⁴ The AI Act between Digital and Sectoral Regulations, Bertelsmann Stiftung, Gütersloh December 2024, p.15

²³⁵ <https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai>

²³⁶ https://www.europarl.europa.eu/charter/pdf/text_en.pdf

²³⁷ [https://www.europarl.europa.eu/RegData/etudes/fiches_techniques/2013/010106/04A_FT\(2013\)010106_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/fiches_techniques/2013/010106/04A_FT(2013)010106_EN.pdf)

*Pursuant to Article 51 (1) of the Charter, the institutions, bodies, offices and agencies of the Union must respect the rights as embodied in the Charter, EU Member States must do so when they are implementing Union Law. This applies equally to AI as to any other field*²³⁸.

This means that when using an AI system, a wide range of Fundamental Rights as enshrined in the Charter should be protected. Using AI systems involves a wide range of fundamental rights including privacy, data protection and non-discrimination²³⁹. In addition, as mentioned in recital 48 of the AI Act, the extent of the unfavorable effect caused by the AI system on the fundamental rights protected by the Charter is very closely related to the classification of an AI system as high risk. *"Those rights include the right to human dignity, respect for private and family life, protection of personal data, freedom of expression and information, freedom of assembly and of association, the right to non-discrimination, the right to education, consumer protection, workers' rights, the rights of persons with disabilities, gender equality, intellectual property rights, the right to an effective remedy and to a fair trial, the right of defense and the presumption of innocence, and the right to good administration"*²⁴⁰. Moreover, as also mentioned in Recital 48 of the AI Act, the fundamental right to a high level of environmental protection also mentioned in the Charter should be taken under consideration when evaluating the seriousness of a damage that an AI system can cause with regards to health and safety of ²⁴¹~~[OBJ]~~.

2.5.1.4 The Council of Europe Framework Convention on Artificial Intelligence

The Council of Europe Framework Convention on Artificial Intelligence²⁴² (hereinafter "the Convention") adopted on 17 May 2024 and opened for signature on 5 September 2024, is the first ever internationally binding treaty in this field, adding on the already existent international standards on human rights, democracy and the rule of law and targets at filling any gaps that could originate from fast technological advances²⁴³. The Framework Convention provides a global uniform legal framework in order to apply to existing international and domestic obligations of each party in the sphere of human rights, democracy and rule of law of each party and intends to secure the consistency of the activities within the lifecycle of artificial intelligence systems with human rights, democracy, and the rule of law²⁴⁴.

As mentioned in article 3 of the Convention defining the scope, each party shall apply the Convention to the activities of the lifecycle of artificial intelligence systems undertaken by public authorities or private actors acting on their behalf and address the risks and impacts arising from activities by private actors not covered as abovementioned. General obligations imposed to parties of the Convention refer to the adoption or maintenance of measures to ensure that artificial intelligence systems are consistent

²³⁸ Getting the future right- Artificial Intelligence and Fundamental Rights, European Union Agency for Fundamental rights, 2021 p.5

²³⁹ Getting the future right- Artificial Intelligence and Fundamental Rights, European Union Agency for Fundamental rights, 2021 p.5

²⁴⁰ Recital 48 of the AI Act

²⁴¹ Recital 48 of the AI Act

²⁴² <https://rm.coe.int/1680afae3c>

²⁴³ <https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence>

²⁴⁴ Explanatory Report to the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, <https://rm.coe.int/1680afae67>

with the protection of human rights²⁴⁵, do not undermine the integrity of democratic processes²⁴⁶, and principles laid down refer to the respect human dignity and individual autonomy²⁴⁷, ensure transparency and oversight²⁴⁸, ensure accountability and responsibility for adverse impacts on human rights²⁴⁹, respect equality and prohibit discrimination²⁵⁰, protect privacy rights of individuals and their personal data²⁵¹, promote reliability of artificial intelligence systems and trust in their outputs²⁵² while parties are called upon to enable establishment of testing environments for artificial intelligence systems under the supervision of their competent authorities²⁵³ and finally take or maintain measures for the identification, assessment and mitigation of risks posed by artificial intelligence systems.

2.5.1.5 *Artificial Intelligence Act*

The Artificial Intelligence Act (hereinafter “AI Act”)²⁵⁴ is the first ever legal framework on AI, addressing the risks related to AI²⁵⁵, including clear directions either by establishing limits to the systems or by defining what should be prohibited²⁵⁶.

Through its implementation, the EU aims to develop a comprehensive framework and ecosystem to enable citizens to make use of the benefits of artificial intelligence while at the same time reducing its risks at EU – wide level²⁵⁷.

The AI Act can be considered a horizontal piece of legislation laying down future proof harmonized rules for the creation, placement on the market and use of AI applications throughout their life cycle in the EU. As mentioned by the explanatory memorandum of the proposal of AI Act, the latter presented a balanced and proportionate regulatory approach to AI that is limited to the minimum necessary requirements to deal with the risks and problems linked to AI, without exceedingly constraining or impeding technological development or otherwise disproportionately augmenting the cost of placing AI solutions on the market²⁵⁸.

²⁴⁵ article 4

²⁴⁶ article 5

²⁴⁷ article 7

²⁴⁸ article 8

²⁴⁹ article 9

²⁵⁰ article 10

²⁵¹ article 11

²⁵² article 12

²⁵³ article 13

²⁵⁴ <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>

²⁵⁵ <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>

²⁵⁶ Can algorithm be ethical, fair, transparent, judge and administer? book, Crete University Press, 2023, Spyros Tassis, p.85, (title translated from greek)

²⁵⁷ Revolutionizing Law? Exploring experimental regulations and artificial intelligence Sandboxes in the AI Act, Rauhe David, <https://eulawanalysis.blogspot.com/2025/03/revolutionizing-law-exploring.html> cross reference Rachondas, Moral + Machines 1/2021, p. 90

²⁵⁸ Explanatory memorandum of the AI Act available at <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206>, 1.1

In addition, the AI Act follows a risk-based approach, differentiating between risk categories, whereas mentioned in literature, these applicable risk categories are largely determined by the intended use²⁵⁹.

This risk-based classification of AI systems, also focusing on the potential harm to health, safety and fundamental rights probably caused by an AI systems, categorizes different risk levels from minimal to high risk²⁶⁰ (unacceptable, high, limited, minimal²⁶¹), meaning that according to the risk level of artificial intelligence systems, the latter are subject to different rules²⁶².

This categorization system ensures that high-risk AI applications, and more specifically those that have the capacity to impact fundamental rights to a significant degree, are either prohibited or subject to more strict requirements and oversight²⁶³. Certain AI systems not included in the categories of prohibited practices or high-risk AI systems should primarily abide by transparency obligations²⁶⁴.

In other words, the AI Act provides a “top-down risk assessment” and does not provide a unique set of rules for all systems and models, instead it divides those systems and models into different classes where each is subject to its own legal framework²⁶⁵. In addition, the AI Act lays down in article 51 et seq a tiered classification systems for General Purpose AI models ranging from standard models to those AI models with systemic risk, introducing specific criteria for this classification.

In short, the AI Act lays down prohibitions of certain AI practices as mentioned in article 5 (e.g. placing on the market, putting into service or use AI systems that deploys subliminal techniques to materially distort the behavior of a person or group that can cause significant harm, that exploits vulnerabilities of a natural person or group that can cause significant harm, that evaluate or classify natural persons or groups based on their social behavior or personality characteristics (social scoring) leading to detrimental or unfavorable treatment of certain natural persons in social contexts detached from the context in which the data was collected in the first place or unjustified to their social behavior etc.). Prohibited AI practices are not expected to appear in the energy sector, thus at this stage of the project they are regarded as less relevant.

Chapter III of the AI Act regulates high risk AI systems through several provisions. Initiating by setting the classification rules for high risk AI systems [AI systems intended to be used as a safety component of a product or the AI systems is itself a product listed in Annex I of the AI Act and is required to undergo a third party conformity assessment prior to putting into service and more closely linked to the project AI systems listed in Annex III of the AI Act among which critical infrastructure meaning AI systems intended to be used as safety components in the management and operation of critical

²⁵⁹ Hannah Ruschemeier, AI as a challenge for legal regulation – the scope of application of the artificial intelligence act proposal, ERA Forum (2023) 23:361–376

<https://doi.org/10.1007/s12027-022-00725-6>

²⁶⁰ Bird and Bird, European Union Artificial Intelligence Act: a guide,

²⁶¹ <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>

²⁶² Friedl, Paul; Gasiola, Gustavo Gil: *Examining the EU's Artificial Intelligence Act*, VerfBlog, 2024/2/07,

<https://verfassungsblog.de/examining-the-eus-artificial-intelligence-act/>, DOI: [10.59704/789d6ad759d0a40b](https://doi.org/10.59704/789d6ad759d0a40b).

²⁶³ European Union Artificial Intelligence Act: a guide, April 2025, Bird and Bird, p. 3

²⁶⁴ The EU Artificial Intelligence (AI Act), a commentary, Wolters Kluwer, 2025, p. 9, article 50

²⁶⁵ Law & Compliance in AI Security & Data Protection, AI and Data Protection Training Module,

[https://www.dpa.gr/sites/default/files/2025-](https://www.dpa.gr/sites/default/files/2025-03/Law%20%26%20Compliance%20in%20AI%20Security%20%26%20Data%20Protection.pdf)

[03/Law%20%26%20Compliance%20in%20AI%20Security%20%26%20Data%20Protection.pdf](https://www.dpa.gr/sites/default/files/2025-03/Law%20%26%20Compliance%20in%20AI%20Security%20%26%20Data%20Protection.pdf), p. 8

infrastructure or in the supply of gas, heating or electricity²⁶⁶]. Such an example of a high-risk AI system may be an AI system that controls an electrical substation, provides balance to supply and demand on the grid, or improves a power plant's output if its malfunction might endanger safety or service continuity²⁶⁷.

In the same chapter of the AI Act, requirements for high risk AI systems are included among which the establishment, implementation and documentation of a risk management system, adoption of data governance and management practices (e.g. relevant data preparation processing operations, examination in view of probable biases that can affect the health and safety of persons as well as measures to secure that training, validation and testing data sets are relevant, representative and to the best extent possible free of errors²⁶⁸ etc.) the drawing up before the system is placed on the market and kept up to date of technical documentation of a high risk AI system, enabling automatic record keeping over the life time of the system²⁶⁹, securing transparent operation of the system to deployers and provision of instructions for use²⁷⁰, adoption of human oversight measures to prevent risks to health, safety and fundamental rights²⁷¹, development of the system in a way that it can achieve appropriate level of accuracy, robustness and cybersecurity²⁷².

A detailed list of obligations of providers of high-risk AI systems is also included in the AI Act²⁷³ including all requirements above mentioned as well indication of certain information, disposition of a quality management system, maintenance of documentation, log keeping, ensuring that the system undergoes the relevant conformity assessment procedure, drawing up an EU declaration of conformity, affix of the CE marking, compliance with registration obligations, taking the necessary corrective actions when the system is not in compliance with AI Act, demonstration of conformity when requested by a competent authority, compliance with accessibility requirements.

Moreover a list of specific obligations is also laid down for deployers of high risk AI systems among which the taking of appropriate technical and organizational measures to ensure that they use the high risk AI systems in accordance with instructions for use²⁷⁴, assignment of appropriate human oversight, securing that input data is relevant and sufficiently representative when the deployer exercises control over input data, monitoring of the operation of the system based on the instructions of use, keeping of logs automatically generated by the high risk AI system to the degree that such logs are subject to their control, information of workers' representatives before using a high risk AI system at the workplace, registration obligations for deployers that are public authorities, when applicable use of the information provided under article 13 of AI Act to carry out data protection impact assessment, information to data subjects for systems of Annex III that make decisions or assist in making decisions

²⁶⁶ Article 6 of AI Act

²⁶⁷ Jørgensen, B.N.; Ma, Z.G. Regulating AI in the Energy Sector: A Scoping Review of EU Laws, Challenges, and Global Perspectives. *Energies* 2025, 18, 2359. <https://doi.org/10.3390/en18092359>

²⁶⁸ Article 10 of the AI Act

²⁶⁹ Article 11 and 12 of the AI Act

²⁷⁰ Article 13 of the AI Act

²⁷¹ Article 13 of the AI Act

²⁷² Article 15 of the AI Act

²⁷³ Article 16 of the AI Act with reference to specific articles

²⁷⁴ Article 26 of the AI Act

related to natural persons, performance of fundamental rights impact assessment when applicable²⁷⁵. Other obligations apply to other parties such as importers and distributors.

The AI Act also lays down rules for general-purpose AI models and general-purpose AI models that pose systemic risk. Starting from the definition of the general purpose AI model in article 3 (63) of the AI Act and general purpose AI system in article 3 (66) and their categorization to models with systemic risk in article 51, specific obligations are established for providers of general purpose of AI models, such as drawing up and keeping up to date documentation of the model with minimum information for the purpose to be provided upon request to AI office and national competent authorities²⁷⁶, drawing up and keeping up to date and making available information and documentation to providers of AI systems who plan to integrate the GPAI model in the AI system, put in place a policy to comply with copyright and related rights law and drawing up and making publicly available an adequately elaborated summary about the content used for training GPAI model.

In addition to obligations for providers of GPAI models, those of models with systemic risk have additional obligations such as performance of model evaluation, assessment and mitigation of possible systemic risks including their sources, maintenance and report to the AI office and national competent authorities of information about serious incidents and possible corrective measures and ensuring an adequate level of cybersecurity protection²⁷⁷.

As to the rules regarding transparency according to article 50 of the AI Act, providers of AI systems intended to interact directly with natural persons should ensure through the development and design of the system that the natural persons concerned are notified about their interaction with an AI system (i.e. an AI powered customer service agent for utility). Moreover, providers of AI systems, including general-purpose AI systems, generating synthetic audio, image, video or text content, shall make sure the marking of the results of the AI system in a machine-readable format and be able to be distinguished as artificially generated or manipulated and "deployers of an AI system that generates or manipulates image, audio or video content constituting a deep fake, shall disclose that the content has been artificially generated or manipulated". It should be noted that transparency obligations are also applicable to providers of GPAI systems.

With regards to copyright, AI Act introduces compliance obligations for providers of general purpose AI models, requiring them to implement policies that align with the EU copyright laws, including the use of state of the art technologies to identify copyrighted works and comply with the TDM opt-out mechanism established in Digital Single Market Directive (DSMD further analyzed below). Providers of general-purpose AI models are obliged to make public elaborated summaries about the datasets used for training. Rules regarding regulatory sandboxes and testing of high-risk AI systems in real world conditions outside regulatory sandboxes are also relevant and will be analyzed in section 1.5.2 below.

2.5.1.6 General Data Protection Regulation

General Data Protection Regulation (hereinafter "GDPR") includes several terms referring to the Internet (Internet, social networks, website, links etc.) but contains neither the term "Artificial Intelligence" as

²⁷⁵ Article 26 of the AI Act

²⁷⁶ Article 53 of the AI Act

²⁷⁷ Article 55 of the AI Act

such nor terms or related notions such as machine learning or big data²⁷⁸. Still, the fundamental right to the protection of personal data is safeguarded mainly through GDPR and AI Act neither seeks to affect existing law governing the processing of personal data, nor affects the obligations of providers and deployers of AI systems with regards to their role as data controllers or processors deriving from Union or national law on the protection of personal data to the extent that the design, the development or use of AI systems involves the processing of personal data²⁷⁹. As commented by Jørgensen et al, while GDPR does not target to particularly rule AI or energy sector, it has remarkable implications for those AI applications in the energy sector that involve personal data and may be mostly related to demand side and retail market such as smart meter data linked to households, consumer energy usage profiles, billing and payment data.

As such obligations imposed by GDPR including among other data protection by design, implementation of appropriate technical and organizational measures, records of processing activity, written contract between controller and processor when the processing of personal data is carried on behalf of the controller, are applicable. In addition, data subjects continue to enjoy all the rights and guarantees, including the rights related to solely automated individual decision-making including profiling. Moreover, data subjects should be enabled to exercise their rights and other remedies provided for the protection of personal data in European Law. Transparency about the original purpose of the data collection should be secured as data governance and management practice²⁸⁰. As specifically mentioned in the scope of the AI Act, the latter does not affect GDPR and is without prejudice to article 10 (5) and 59 of the AI Act. Indicative provisions of GDPR related to AI applications in energy sector include lawful basis and purpose limitation (for example the need of an acceptable legal basis in case of an AI application processing personal data originating from smart thermostat data to optimize home energy consumption), data minimization and privacy by design (for example anonymization or aggregation of energy data are preferred so that AI models avoid using personal data if not necessary), automated decision-making and profiling (for example in an AI system that arranges in an automatic way a user's tariffs or disconnects service for non-payment of energy bills), security and breach notification obligations when required as well as conducting DPIAs (for example in case of large scale monitoring through smart devices)²⁸¹.

As to further processing of personal data for developing certain AI systems in the public interest in the AI regulatory sandbox (please find further analysis for regulatory sandboxes below) GDPR is also applicable. Based on article 59 of the AI Act, in the AI regulatory sandbox, personal data lawfully collected for other purposes may only be processed for the purpose of developing, training and testing certain AI systems in the sandbox when all the conditions mentioned therein apply. More specifically, as mentioned in recital 140 of the AI Act, AI Act provides "the legal basis for the providers and prospective providers in the AI regulatory sandbox to use personal data collected for other purposes

²⁷⁸ The impact of General Data Protection Regulation (GDPR) on artificial intelligence, Study, Panel for the Future of Science and Technology, EPRS, European Parliamentary Research Service, June 2020, [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU\(2020\)641530_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf). p. 35

²⁷⁹ Recital 10 of AI Act

²⁸⁰ Recital 67 of AI Act

²⁸¹ Jørgensen, B.N.; Ma, Z.G. Regulating AI in the Energy Sector: A Scoping Review of EU Laws, Challenges, and Global Perspectives. *Energies* 2025, 18, 2359. <https://doi.org/10.3390/en18092359>

for developing certain AI systems in the public interest within the AI regulatory sandbox”, only in specified conditions, along with the applicability of certain requirements provided in GDPR such as article 6 (4) and 23 (1). Article 6 (4) of the GDPR permits processing of personal data for a different purpose compared to that for which the personal data have been collected based on Union law which must be a measure that is necessary and proportionate in a democratic society to achieve the objectives referred to in article 23 (1) of the GDPR. Article 23 (1) of the GDPR allows the restriction of rights provided for in articles 12 to 22 and 34 of GDPR as well as article 5, when such restriction respects the essence of fundamental rights and freedom. Thus, in such cases and according to the above-mentioned provisions and conditions, GDPR is also applicable.

2.5.1.7 Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union

The Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union²⁸² (hereinafter “Regulation for the free flow of non-personal data”) aims at ensuring the free flow of (electronic) data other than personal data within the Union. More specifically, as defines in article 1 regarding subject matter, it governs data localization requirements, availability of data to competent authorities and the porting of data to professional users and applies to the processing of electronic data provided as a service to users residing or having an establishment in the Union regardless of the establishment of the service provider or carried out by a person residing or having an establishment in the Union for its own needs.

The Regulation for the free flow of non-personal data prohibits data localization requirement meaning “*any obligation, prohibition, condition, limit or other requirement provided for in the laws, regulations or administrative provisions of a Member State[...]*”²⁸³ *which imposes the processing of data in the territory of a specific Member State or hinders the processing of data in any other Member State*”. The abovementioned Regulation includes rules to secure availability for competent authorities to request, obtain, and access data.

2.5.1.8 Data Act

The Data Act (Regulation EU 2023/2854) which came into force on the 11th of January 2024 and shall apply from 12 September 2025 with some exceptions that will apply later, lays down harmonized rules to enable a fair distribution of the personal and non-personal data by establishing concrete and fair rules regarding the access and use of data in the framework of the European Data Economy²⁸⁴. This regulation provides support in innovative energy services by defining and making clear the rules on data usage and access²⁸⁵.

²⁸² <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018R1807>

²⁸³ full definition in article 3 par. 5

²⁸⁴ https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202302854

²⁸⁵ Jorgensen, B.N.; Ma, Z.G. Regulating AI in the Energy Sector: A Scoping Review of EU Laws, Challenges, and Global Perspectives. *Energies* **2025**, *18*, 2359. <https://doi.org/10.3390/en18092359>

More specifically, the Data Act regulates among others the sharing of product data and related service data to the user of the connected product or related service, the sharing of data-by-data holder to data recipient, the sharing of data-by-data holders to public sector bodies²⁸⁶.

In addition, it facilitates switching between data processing services, includes safeguards to prohibit unlawful access to non-personal data, and inserts rules on the creation of interoperability standards for the access, transfer, and use of data²⁸⁷. Data Act refers to three types of data sharing processes, business to consumer, business to business and business to governments. It introduces among other obligations such as obligation to make product data and related data accessible to the user, enables the user to share data with third parties upon request to a data holder, imposing obligations to the third party receiving the data.

The conditions under which data holders can make data available to data recipients are also laid down²⁸⁸ (e.g. fair and transparent terms, non-discrimination in arrangements for making data available, non-disclosure of trade secrets etc.), as well as procedural issues (compensation, dispute settlement) and technical measures to protect from unauthorized data access²⁸⁹. Rules regarding the prohibitions for the inclusion of unfair contract terms concerning access, use and liability imposed unilaterally on another business are also included. The making available of data to public sector bodies, Commission etc. in an exceptional need in the public interest is also regulated as well as the obligation of public sector bodies. Furthermore, rules regarding switching between data processing services are included especially with regards to facilitate switching and remove obstacles²⁹⁰ as well as the necessity for clear written contract ruling switching and minimum content and necessary information from providers of data processing services²⁹¹. Technical aspects such as the adoption of reasonable measures, availability of interfaces, and compatibility with common specifications based on open interoperability specifications are laid down. In addition, issues regarding interoperability of data such as essential requirements and interoperability of data processing services objectives are laid down²⁹².

2.5.1.9 Directive on privacy and electronic communications/ e-privacy

The e-privacy Directive (Directive 2002/58/EC)²⁹³ has been amended by Directive 2009/136²⁹⁴ and subject to its transposition to the legislation of member states, protects private life and confidentiality of communications, including the way of providing conditions for any storing of personal and non-personal data in and access from terminal equipment.

The e-privacy Directive along with GDPR provide the basis for sustainable and responsible data processing for data sets with both personal and non-personal data²⁹⁵. More specifically, this Directive

²⁸⁶ article 1

²⁸⁷ article 1

²⁸⁸ article 8

²⁸⁹ articles 9-11

²⁹⁰ article 23

²⁹¹ article 25

²⁹² article 33

²⁹³ <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32002L0058>

²⁹⁴ <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32009L0136>

²⁹⁵ Recital 10 of AI Act

includes, among others, rules regarding the lawful use of cookies as well as the processing of data falling within its scope and unsolicited communications. The e–privacy Directive continues to apply to personal data processed in relation to the rights and obligations provided in the AI Act.

2.5.1.10 NIS 2 Directive

The NIS 2 (Directive EU 2022/2555) Directive²⁹⁶ is the successor of the NIS Directive (Directive EU 2016/1148) and is considered as “a cornerstone of the European Union’s efforts to ensure a high common level of cybersecurity across all member states”²⁹⁷. As aptly commented by Jorgensen N et al, although NIS 2 does not directly and specifically target AI, still it reinforces the obligations across sectors, indirectly implementing higher security standards for AI systems in energy²⁹⁸.

Due to its nature as a Directive, NIS 2 is not directly applicable to Member States but has to be transposed in national legislation. As mentioned in recital 6 of NIS 2, with the repeal of NIS Directive through NIS 2, the later has extended the scope of application to a larger part of the economy to provide a comprehensive coverage of the sectors and the services of vital importance to key societal and economic activities in the internal market.

The energy sector and specifically the subsectors specified in Annex I of the Directive (electricity, district heating and cooling, oil, gas and hydrogen) are included in high criticality sectors which, subject to the criteria defined in article 2 of the Directive, fall within its scope. As also mentioned by Jorgensen et al, NIS 2’s impact on AI solutions is not direct but it is important since “*it forces energy sector organizations to pay robust attention to the security and resilience of all digital systems including AI*”²⁹⁹. In addition, it should be noted that an important intersection between NIS 2 and the AI Act is located at the point where high risks AI systems should be resilient against manipulation and deal with adversarial threats³⁰⁰.

Entities falling within the scope of the Directive bear specific obligations with regards to governance (article 20 of the Directive) such as approval of cybersecurity risk management measures and appropriate training of members of the management bodies and employees³⁰¹. Moreover covered entities should take appropriate and proportionate technical and operational measures defined in article 21 of the Directive, to manage the risks posed to security of network and information systems which those entities use for their operations or for the provision of their services, and to prevent or minimize the impact of incidents on recipients of their services and on other services.

²⁹⁶ <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32022L2555>

²⁹⁷ <https://www.enisa.europa.eu/topics/awareness-and-cyber-hygiene/raising-awareness-campaigns/network-and-information-systems>

²⁹⁸ Jørgensen, B.N.; Ma, Z.G. Regulating AI in the Energy Sector: A Scoping Review of EU Laws, Challenges, and Global Perspectives. *Energies* 2025, 18, 2359. <https://doi.org/10.3390/en18092359>

²⁹⁹ Jørgensen, B.N.; Ma, Z.G. Regulating AI in the Energy Sector: A Scoping Review of EU Laws, Challenges, and Global Perspectives. *Energies* 2025, 18, 2359. <https://doi.org/10.3390/en18092359>

³⁰⁰ Jørgensen, B.N.; Ma, Z.G. Regulating AI in the Energy Sector: A Scoping Review of EU Laws, Challenges, and Global Perspectives. *Energies* 2025, 18, 2359. <https://doi.org/10.3390/en18092359>

³⁰¹ Article 20 of NIS 2 Directive

An indicative example of NIS 2 implications with regards to AI applications in the energy sector is in case a distribution system operator makes use of an AI to autonomously control voltage or detect outages; this AI system and its data flows should abide by security requirements³⁰²

2.5.1.11 Cybersecurity Act

This Regulation (EU 2019/881) strengthens the role of ENISA and sets up a European Framework for cybersecurity certification for ICT products, services and processes in the Union. Although the Act does not lay down direct obligations, it provides the tools and trust mechanisms that can contribute to AI compliance and risk management in the energy sector³⁰³. According to article 42 of the AI Act, high-risk AI systems that have been certified or for which a statement of conformity has been issued under a cybersecurity scheme pursuant to Regulation (EU) 2019/881³⁰⁴ and the references of which have been published in the Official Journal of the European Union shall be presumed to abide by the requirements of cybersecurity set out in Article 15 of AI Act in so far as the cybersecurity certificate or statement of conformity or parts thereof cover those requirements. Cybersecurity Act does not directly impose obligations but consists of “*the foundation of cybersecurity ecosystems in which AI driven energy solutions operate*” while additionally, it lays down key tools for trust³⁰⁵.

2.5.1.12 Cyber Resilience Act (CRA)

The Cyber Resilience Act (Regulation (EU) 2024/2847)³⁰⁶ fortifies the security of AI and digital components in the energy systems by laying down rules for the products themselves³⁰⁷. Its relevance to the energy sector is linked to the fact that many energy-related systems depend on digital products which insert or support AI³⁰⁸. In addition, according to article 15 of the AI Act, high risk AI systems shall be designed and developed in such a way that ensures appropriate levels of accuracy, robustness and cybersecurity.

Technical and organizational measures should be taken to ensure that high Risk AI systems should be resilient as possible regarding errors, faults or inconsistencies that may take place within the system of the environment in which the system operates and also resilient against attempts by unauthorized third

³⁰² Jørgensen, B.N.; Ma, Z.G. Regulating AI in the Energy Sector: A Scoping Review of EU Laws, Challenges, and Global Perspectives. *Energies* 2025, 18, 2359. <https://doi.org/10.3390/en18092359>

³⁰³ Jørgensen, B.N.; Ma, Z.G. Regulating AI in the Energy Sector: A Scoping Review of EU Laws, Challenges, and Global Perspectives. *Energies* 2025, 18, 2359. <https://doi.org/10.3390/en18092359>

³⁰⁴ <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019R0881>

³⁰⁵ Jørgensen, B.N.; Ma, Z.G. Regulating AI in the Energy Sector: A Scoping Review of EU Laws, Challenges, and Global Perspectives. *Energies* 2025, 18, 2359. <https://doi.org/10.3390/en18092359>

³⁰⁶ https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202402847

³⁰⁷ Jørgensen, B.N.; Ma, Z.G. Regulating AI in the Energy Sector: A Scoping Review of EU Laws, Challenges, and Global Perspectives. *Energies* 2025, 18, 2359. <https://doi.org/10.3390/en18092359>

³⁰⁸ Jørgensen, B.N.; Ma, Z.G. Regulating AI in the Energy Sector: A Scoping Review of EU Laws, Challenges, and Global Perspectives. *Energies* 2025, 18, 2359. <https://doi.org/10.3390/en18092359>

parties to modify their use, outputs or performance by taking advantage of system's vulnerabilities³⁰⁹. Compliance with cybersecurity measures against unauthorized system alterations can be shown through the EU Cyber Resilience Act for those cases where systems are subject to the Cyber Resilience Act³¹⁰.

Specific requirements, for products with digital elements which fall within the Scope of Cyber Resilience Act, and which are classified as high-risk AI systems to be deemed compliant with cybersecurity requirements of Article 15 of the AI Act, are laid down in article 12 of the Cyber Resilience Act. Most of the Regulation's provisions shall apply from 11 December 2027.

2.5.1.13 Digital Single Market Directive

General purpose AI models and more specifically large generative AI models which have the capacity to generate text, images and other content, present important challenges to creators and the way their creative content is created, distributed, used, and consumed³¹¹.

Since it can be said that for the development and training of such models it is necessary to acquire access to significant volume of text and other data, text and mining techniques are probable to be used extensively to extract and analyze content which may be protected by copyright and related rights. According to copyright legislation, any use of copyrighted content is lawful following the authorization by the rightsholder unless copyright exceptions and limitations apply.

The DSMD (Directive 2019/790)³¹² introduced exceptions for "text and mining" (TDM purposes)³¹³ in article 4 (3) meaning the exception or limitation for reproductions and extractions of lawfully accessible work and other subject matter for the purposes of text and mining on condition that the use of works and other subject matter has not been expressly reserved by their rightsholder in an appropriate manner. Based on the provisions of AI Act, GPAI providers should comply with copyright law and the opt-out exception of the Copyright Directive, which permits TDM if rightsholders do not express their refusal and make public a sufficiently detailed summary proving information on the content used for training³¹⁴.

In addition, in article 3, DSMD has also enacted text and data mining exceptions to the reproduction rights for the purpose of scientific research by research and cultural institutions.

³⁰⁹ article 15 of AI Act

³¹⁰ Bird and Bird, European Union Artificial Intelligence Act: A guide, 7 April 2025, p. 29

³¹¹ Recital 105 of the AI Act

³¹² <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L0790>

³¹³ AI and copyright: The training of general purpose AI, Tristan Marcelin with Filippo Caseti, April 2025, [https://www.europarl.europa.eu/RegData/etudes/ATAG/2025/769585/EPRS_ATAG\(2025\)769585_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/ATAG/2025/769585/EPRS_ATAG(2025)769585_EN.pdf)

³¹⁴ AI and copyright: The training of general purpose AI, Tristan Marcelin with Filippo Caseti, April 2025, [https://www.europarl.europa.eu/RegData/etudes/ATAG/2025/769585/EPRS_ATAG\(2025\)769585_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/ATAG/2025/769585/EPRS_ATAG(2025)769585_EN.pdf)

2.5.1.14 *Information society Directive*

In the AI ecosystem vast datasets serve as the basis for generative AI systems to learn, adapt and make decisions³¹⁵.

However, the use of such datasets for training purposes raises question regarding copyright and ownership. When training generative AIs, various techniques are used including text and data mining³¹⁶. Given that under Information Society Directive³¹⁷, the right of reproduction is laid down, meaning the right of the creator to authorise or prohibit direct or indirect temporary or permanent reproduction³¹⁸ and exceptions to it are also defined, many copies that are made in the process of text and data mining can be generally assessed as act of reproduction, thus can raise copyright infringement issues.

The exception of temporary reproductions of article 5 of Information Society Directive, are excepted from the authorisation of reproduction right when these acts are transient or incidental and an integral and essential part of a technological process whose sole purpose is to enable a transmission in a network or lawful use of a work to be made and which have no independent economic significance.

In addition, the exceptions of DSMD for data mining for the training of AI systems are subject to article 5(5) of the Information Society Directive, which states that those limitations or exceptions will be only applied for those "certain special cases which do not conflict with a normal exploitation of the work or other subject – matters and do not unreasonably prejudice the legitimate interests of the rightsholder³¹⁹".

2.5.1.15 *Directive 96/9 on the legal protection of databases*

Subject to national transposition, Directive 96/9³²⁰ on the legal protection of databases lays down the protection of databases by copyright.

More specifically, based on article 3 of the abovementioned Directive, databases which "by reason of the selection or arrangement of their contents" are considered their own intellectual creation shall be protected by copyright.

In this Directive restricted acts which require author's authorization include temporary or permanent reproduction by any means, and in any form in whole or in part, translation, adaptation, arrangement and any other alteration as well as any form of distribution to the public of the database as well as any communication, display or performance to the public. The above-mentioned Directive is relevant in

³¹⁵ Hans Graux, Pieter Gryffroy, Magdalena Gad- Nowak, Liesa Boghaert, Data Europe EU, The role of artificial intelligence in the processing and generating new data, doi: 10.2830/412108

³¹⁶ Artificial intelligence and copyright: use of generative AI tools to develop new content, https://intellectual-property-helpdesk.ec.europa.eu/news-events/news/artificial-intelligence-and-copyright-use-generative-ai-tools-develop-new-content-2024-07-16-0_en

³¹⁷ Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the information society, <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32001L0029>

³¹⁸ article 2

³¹⁹ Artificial intelligence and copyright: use of generative AI tools to develop new content, https://intellectual-property-helpdesk.ec.europa.eu/news-events/news/artificial-intelligence-and-copyright-use-generative-ai-tools-develop-new-content-2024-07-16-0_en

³²⁰ Directive 96/9 on the legal protection of databases, <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31996L0009>

cases where data scraping may involve extracting substantial parts of a database protected by copyright, thus resulting in copyright violation³²¹.

2.5.1.16 Digital Services Act

The Regulation (EU) 2022/2065 lays down rules for online platforms and recommender systems, including AI based content moderation³²². AI systems may be provided as intermediary services or parts thereof within the meaning of the Digital Services Act³²³ in which case the AI Act should be without prejudice to the liability of intermediary services regulating provisions³²⁴.

The Digital Services Act aims to set harmonized rules for a secure, foreseeable and trusted online environment that promotes innovation and protects consumers.

In the Digital Services Act, among others the liability of providers of intermediary services according to types defined is laid down, and a system of accountability and transparency for intermediary services is introduced. Special rules are included for very large online platforms and very large online search engines.

The AI Act applies without prejudice to the liability provisions for providers of intermediary services set out in the Digital Services Act. Still obligations derived from DSA will not be examined in this project.

2.5.1.17 European Declaration on Digital Rights and Principles for the Digital Decade

The European Declaration on Digital Rights and Principles³²⁵ "*spells out shared political intentions and commitments and recalls the most relevant rights in the context of digital transformation*"³²⁶ and also should provide guidance to policy makers when considering the digital transformation, while at the same time should function as a reference point for relevant actors and businesses when creating and using new technologies. The parties aim at promoting digital transformation by putting people at the center and reaffirming universal human rights and benefit of all individuals, businesses and society. As mentioned by Patrick Van Eecke & Bartholomäus Regenhardt, "*AI reflects the declaration's focus on fundamental rights, ensuring AI applications are developed and used in a way that respects these rights*"³²⁷.

In the Declaration, the following have been declared:

- A) People are put at the center of digital transformation
- B) Digital transformation should contribute to solidarity and inclusion; everyone should have access to high – speed digital connectivity. Everyone should be able to develop all basic and advanced digital skills and is entitled to fair, just and safe working conditions in the digital environment

³²¹ OECD, Intellectual property issues in artificial intelligence trained on scraped data, OECD artificial intelligence papers, February 2025, No 33

³²² White & Case, EU AI Act Handbook, May 2025, p. 102

³²³ <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32022R2065>

³²⁴ Recital 11 of the AI Act

³²⁵ <https://digital-strategy.ec.europa.eu/en/library/european-declaration-digital-rights-and-principles>

³²⁶ <https://digital-strategy.ec.europa.eu/en/library/european-declaration-digital-rights-and-principles>, paragraph 7

³²⁷ The EU Artificial Intelligence (AI) Act, a commentary, Wolters Kluwer, 2025, p. 13

- C) Freedom of choice should be ensured and everyone should be able to profit from the benefits of algorithmic and artificial intelligence systems including by forming their personal and informed choices in the digital environment and a fair digital environment should be secure both through free choice of online services that can be used and possibility to compete fairly and innovate on the digital environment
- D) Participation in the digital public sphere should be enabled and supported
- E) Safety, security and empowerment should be further promoted through a protected, safe and secure digital environment and protection of privacy and protection of children and young people in the digital environment
- F) Sustainability should be supported and promoted.

2.5.1.18 Approval of the content of the draft Communication from the Commission - Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act), European Commission, 6-2-2025.

The Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act) (hereinafter "Guidelines")³²⁸ provide guidance for the determination as to whether a system constitutes an AI system within the meaning of the AI Act, so as to enable further assessment of the application of the AI Act as to the fulfillment of the criterion of the AI system. As clarified in the Guidelines, the latter are not binding, and the authoritative interpretation of the AI Act may ultimately only be provided by the Court Justice of the European Union³²⁹.

The analysis of the AI system's elements provided in the Guidelines is based on the definition of AI system as laid down in Article 3 (1) of the AI Act, which has as follows: "*AI system' means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments*".

Guidelines analyze each aspect of AI system definition abovementioned, according to which the seven main elements of an AI system are³³⁰:

- [1] a machine-based system.
- [2] that is designed to operate with varying levels of autonomy.
- [3] that may exhibit adaptiveness after deployment.
- [4] and that, for explicit or implicit objectives.
- [5] infers, from the input it receives, how to generate outputs
- [6] such as predictions, content, recommendations, or decisions

³²⁸ <https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act>

³²⁹ ANNEX to the Communication to the Commission Approval of the content of the draft Communication from the Commission - Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act), European Commission, 6-2-2025, par. 7

³³⁰ ANNEX to the Communication to the Commission Approval of the content of the draft Communication from the Commission - Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act), European Commission, 6-2-2025, par. 9

[7] that can influence physical or virtual environments.

More specifically, the element of “machine based” means that AI systems are developed with and run in machines, meaning they require machines to enable their functioning underlining that they must be computationally driven and based on machine operations³³¹, thus covering a variety of computational systems. Moreover, as to the element of autonomy, based on recital 12 of the AI Act “varying levels of autonomy” means that AI Systems are designed to function “with some degree of independence of actions from human involvement” and of the capabilities to operate without human intervention³³², rendering the level of autonomy an important condition for the determination of whether a system qualifies as an AI system³³³.

As to the element of adaptiveness that a system may exhibit and based on recital 12 of the AI Act “adaptiveness” relates to self-learning capabilities, enabling the behavior of the system to change while in use. In addition, with regards to explicit or implicit objectives that AI system disposes, explicit are clearly stated objectives that are directly encoded by the developer into the system, while implicit “*refer to goals that are not explicitly stated but may be deduced from the behavior or underlying assumptions of the system*”³³⁴.

As to the ability to infer, which functions as a condition to distinguish AI systems from other types of systems³³⁵, an AI system must be able to infer, from the input it receives, how to generate outputs. Additionally, as to the system’s capacity to generate the outputs that can influence physical or virtual environments, such as predictions, content, recommendations or decisions, this is considered fundamental for the differentiation of an AI system with other systems/forms or software, while the type of output is important to understanding the functionality and impact of an AI System³³⁶. Finally,

³³¹ ANNEX to the Communication to the Commission Approval of the content of the draft Communication from the Commission - Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act), European Commission, 6-2-2025, par. 12

³³² ANNEX to the Communication to the Commission Approval of the content of the draft Communication from the Commission - Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act), European Commission, 6-2-2025, par. 14

³³³ ANNEX to the Communication to the Commission Approval of the content of the draft Communication from the Commission - Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act), European Commission, 6-2-2025, par. 20

³³⁴ ANNEX to the Communication to the Commission Approval of the content of the draft Communication from the Commission - Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act), European Commission, 6-2-2025, par. 24

³³⁵ ANNEX to the Communication to the Commission Approval of the content of the draft Communication from the Commission - Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act), European Commission, 6-2-2025, par. 26

³³⁶ ANNEX to the Communication to the Commission Approval of the content of the draft Communication from the Commission - Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act), European Commission, 6-2-2025, par. 52

as to the interaction with the environment, this means that AI systems can actively impact the environments in which they are deployed³³⁷.

2.5.1.19 Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, EDPB

The Opinion 28/2024 of EDPB on certain data protection aspects related to the processing of personal data in the context of AI models, was issued following a request filed before European Data Protection Board, by the Irish Supervisory Authority regarding fundamental data protection issues raised by the development and deployment of AI models. The questions submitted relate to the following issues³³⁸.

When and how an AI Model can be considered “anonymous”. Taking under consideration that for this question, EDPB focused on the case of AI systems that their design does not aim at providing personal data related to the training data³³⁹, EDPB considered that “*considers that AI models trained on personal data cannot, in all cases, be considered anonymous. Instead, the determination of whether an AI model is anonymous should be assessed, based on specific criteria, on a case-by-case basis*”³⁴⁰. In addition EDPB considered that “*for an AI Model to be considered anonymous, using reasonable means, both (i) the likelihood of direct (including probabilistic) extraction of personal data regarding individuals whose personal data were used to train the model; as well as (ii) the likelihood of obtaining, intentionally or not such personal data from enquiries should be insignificant for any data subject*”³⁴¹. The likelihood should be assessed considering “all means reasonably likely to be used” by the controller or another person and should also take into consideration unintended (re) use or disclosure of the model. Moreover, elements for the evaluation of the likelihood of identifications were mentioned.

How can controllers demonstrate the appropriateness of legitimate interest as legal basis in the development and deployment phases. For the supervising authorities to examine the applicability of the legal basis of the legitimate interest (art. 6 (1)f of GDPR, the three steps test should be conducted, meaning that the following cumulative conditions should be met:

“(i) the pursuit of a legitimate interest by the controller or by a third party.

(ii) the processing is necessary to pursue the legitimate interest; and

(iii) the legitimate interest is not overridden by the interests or fundamental rights and freedoms of the data subjects”³⁴².

³³⁷ ANNEX to the Communication to the Commission Approval of the content of the draft Communication from the Commission - Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act), European Commission, 6-2-2025, par. 59

³³⁸ Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, EDPB, par. 13

³³⁹ Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, EDPB, par. 30

³⁴⁰ Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, EDPB, par. 33

³⁴¹ Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, EDPB, par. 43

³⁴² Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, EDPB, par. 43

Specific guidance is provided by EDPB as to parameters to be taken into consideration while conducting the abovementioned three steps test within the challenging environment of AI.

Whether the unlawful processing of personal data in the development phase affects the legitimacy of the subsequent processing or operation of the AI model. By analyzing three different scenarios, EDPB has noted that a case-by-case analysis should take place especially to assess whether the development and deployment phases serve different purposes and whether the absence of legal basis for the initial processing operation affects the lawfulness of the subsequent processing³⁴³.

In addition, it should be examined whether an evaluation has taken place by the controller deploying the model to demonstrate compliance with the principle of lawfulness, fairness and transparency and in accordance with the appropriate legal basis, to ascertain whether an AI model was not developed by unlawfully processing personal data³⁴⁴. Finally, EDPB has laid down that when controllers subsequently process personal data collected during the deployment phase, after the model has been anonymized, the GDPR should apply in relation to these processing activities and such cases the lawfulness of the processing taking place in the deployment phase should not affect the unlawfulness of the processing at the initial phase³⁴⁵.

2.5.1.20 Methodology for the risk and impact assessment of artificial intelligence systems from the point of view of human rights, democracy and the rule of law (HUDERIA methodology), Council of Europe, Committee on Artificial Intelligence (CAI), 28-11-2024

HUDERIA Methodology³⁴⁶ is a tool adopted by the CAI of Council of Europe on 28-11-2024 with the aim to provide guidance and a structured approach for the conduct of risk and impact assessments for AI systems with regards to human rights, democracy and the rule of law³⁴⁷. It will be complemented in 2025 by the Huderia Model. Huderia methodology is a standalone non-legally binding guidance with no legal effect allowing adaptation to different contexts³⁴⁸. Huderia Methodology includes the following four elements:

- A) The Context-Based Analysis (COBRA): an approach to the required information collection and mapping to understand and spot the risks the AI system may present to human rights, democracy and the rule of law.
- B) The Stakeholder Engagement Process (SEP): an approach to enable involvement of the appropriate stakeholders to collect the necessary information.
- C) The Risk and Impacts Assessment (RIA): steps on the assessment of the risks and impacts to the abovementioned sectors.
- D) The Mitigation Plan: possible steps to mitigation and remedies.

³⁴³ Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, EDPB, par. 122

³⁴⁴ Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, EDPB, par. 129

³⁴⁵ Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, EDPB, par. 135

³⁴⁶ <https://rm.coe.int/cai-2024-16rev2-methodology-for-the-risk-and-impact-assessment-of-arti/1680b2a09f>

³⁴⁷ <https://www.coe.int/en/web/portal/-/huderia-new-tool-to-assess-the-impact-of-ai-systems-on-human-rights>

³⁴⁸ <https://rm.coe.int/cai-2024-16rev2-methodology-for-the-risk-and-impact-assessment-of-arti/1680b2a09f>

2.5.1.21 *Guidelines on prohibited artificial practices established by Regulation (EU) 2024/1689 (AI Act)*

These non-binding Guidelines on prohibited artificial intelligence practices³⁴⁹ (hereinafter Guidelines on prohibited practices) provide legal clarity and insights regarding article 5 of the AI Act.

They repeat AI's Act risk-based approach classification of AI systems into four different risk categories namely unacceptable risk, high risk, transparency risk and minimal or no risk. Guidelines provide a comprehensive analysis of all prohibited practices according to AI Act, meaning harmful manipulation and deception, harmful exploitation of vulnerabilities, social scoring, individual criminal offence risk assessment and prediction, untargeted scraping to develop facial recognition databases, emotional recognition, biometric categorization and real time remote biometric identification.

2.5.2 **Regulatory sandboxes and experimental frameworks**

2.5.2.1 *Background*

Regulatory sandboxes are not an innovative notion, but their application to the AI sector is³⁵⁰. Their origin can be found in promoting innovation in the financial services sector³⁵¹.

In years before the publication of AI Act on 2024, there was no established legal definition for regulatory sandboxes, still bearing as key characteristics to distinguish them from other forms of experimentation regulatory learning, structured approach to testing innovation, controlled real word environment under supervision of one or more competent authorities, explicit link with legislation through a legal base, temporary derogations and exemptions from those part of the legislation that relate to a specific sandbox and the use of appropriate safeguards.

Regulatory sandboxes generally refer to regulatory tools enabling businesses to test and experiment with new and innovative products, services or businesses under supervision of a regulator for a limited period³⁵².

Regulatory sandboxes enable experimentation while guiding regulation toward embracing technologies³⁵³. The JRC in time before the publication of AI Act, defined three characteristics of regulatory experimentation and more specifically that: a) they are introduced for the assistance of innovative solutions that require previous live testing to gather additional factual evidence of their risks and benefits, b) they imply the possibility of providing derogations from the current regulatory

³⁴⁹ <https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act>

³⁵⁰ The EU Artificial Intelligence (AI) Act, a commentary, Wolters Kluwer, 2025, p. 886

³⁵¹ Buocz et al, LAW, INNOVATION AND TECHNOLOGY 2023, VOL. 15, NO. 2, 357–389
<https://doi.org/10.1080/17579961.2023.2245678>, p. 362

³⁵² Parliament 2022: Parliament, Briefing: Artificial intelligence act and regulatory sandboxes,
[https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733544/EPRS_BRI\(2022\)733544_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733544/EPRS_BRI(2022)733544_EN.pdf)

³⁵³ World Bank 2020: World Bank, Global experiences from Regulatory Sandboxes, Fintech Note No 8, p. IX

framework where the latter can be interpreted as an obstacle to the feasibility of the solution and c) they are launched to support regulatory learning³⁵⁴.

Ranchordás also earlier to the publication of AI Act on 2021, has mentioned regulatory sandbox as a new type of experimental legal regime, which is designed to enable market actors to take advantage of less burdensome regulatory conditions than those established by law³⁵⁵. Still as noted by the Basel Committee on Banking Supervision (BCBS) in 2019, sandboxes may imply the use of legally provided discretions by the relevant supervisor however firms participating are obliged to abide by all relevant rules that apply to their activity³⁵⁶.

An important indication showing the expansion of regulatory sandboxes is the introduction of specific provisions in the AI Act³⁵⁷. The AI Act consists of the first horizontal regulatory regime for sandboxes on the Union level³⁵⁸. The Commission stated that the AI Act targets to provide a uniform legal framework for AI, thus promoting innovation through legal certainty³⁵⁹.

In the AI Act, the objectives that regulatory sandboxes intend to serve are fostering innovation and facilitating development of an AI ecosystem, ensuring compliance of the innovative AI systems with AI Act, enhancing legal certainty for innovators and competent authorities oversight and understanding of the opportunities, facilitating regulatory learning for authorities and undertakings, supporting cooperation and sharing of best practices, accelerating access to markets and contributing to evidence based regulatory learning³⁶⁰.

According to a 2020 World Bank study, 57 counties have developed 73 fintech sandboxes³⁶¹. Additional examples of regulatory sandboxes (the list is indicative and not exhaustive) include Japan's regulatory sandbox launched in June 2018 to promote fintech³⁶², Denmark's regulatory privacy sandbox introduced by Danish Data Protection Authority (Datatilsynet) to stimulate privacy enhancing innovation and

³⁵⁴ Gangale, F., Mengolini, A.M., Covrig, L., Chondrogiannis, S. and Shortall, R., Making energy regulation fit for purpose. State of play of regulatory experimentation in the EU, EUR 31438 EN, Publications Office of the European Union, Luxembourg, 2023, ISBN 978-92-68-00415-9, doi:10.2760/32253, JRC132259.

³⁵⁵ Sofia Ranchordás, Experimental lawmaking in the EU: Regulatory Sandboxes, October 2021, University of Groningen Faculty of Law Research Paper Series, No. 12/2021, p. 3

³⁵⁶ Report: Regulatory sandboxes and innovation hubs, European Securities and Markets Authority, 2019 https://www.esma.europa.eu/sites/default/files/library/jc_2018_74_joint_report_on_regulatory_sandboxes_and_innovation_hubs.pdf

³⁵⁷ Ranchordas, Sofia and Vinci, Valeria, Regulatory Sandboxes and Innovation-friendly Regulation : Between Collaboration and Capture (January 16, 2024). Sofia Ranchordas & Valeria Vinci, Regulatory Sandboxes and Innovation-friendly Regulation: Between Collaboration and Capture, Italian Journal of Public Law, Vol. 1/2024, Forthcoming, Tilburg Law School Research Paper, Available at SSRN: <https://ssrn.com/abstract=4696442> or <http://dx.doi.org/10.2139/ssrn.4696442>

³⁵⁸ Hannah Ruschemeier, Thinking Outside the Box? Regulatory Sandboxes as a Tool for AI Regulation 2024-04-07 Pre-print, under review, p.8

³⁵⁹ Buocz et al, LAW, INNOVATION AND TECHNOLOGY 2023, VOL. 15, NO. 2, 357–389 <https://doi.org/10.1080/17579961.2023.2245678>, p. 366

³⁶⁰ Recital 139 and article 57 par. 9 of the AI Act

³⁶¹ World Bank 2020: World Bank, Global experiences from Regulatory Sandboxes, Fintech Note No 8, p. IX

³⁶² <https://www.iflr.com/article/2a63733ixsvbckrrhity/japan-sandbox-reform>

digitization³⁶³, ICO's regulator sandbox to support organizations who are developing products and services which utilize personal data in innovative and safe ways³⁶⁴ and Spain's regulatory sandbox on Artificial Intelligence³⁶⁵ presented on 2022 and aimed to bring competent authorities close to companies that developed AI to define best practices that will guide the implementation of the Artificial Intelligence Act. In April 2025, twelve projects were selected to participate in Spanish Regulatory Sandbox. Other cases of regulatory sandboxes include a data protection sandbox program operated by the national data protection authority (CNIL) not specifically designed as AI Act regulatory sandbox but in its third edition it is offering a new sandbox to support three projects using artificial intelligence for the benefit of public services³⁶⁶ and Luxembourg's regulatory sandbox introduced in May 2024 for testing and understanding the legal implications of new technologies and new uses of data especially in the field of artificial intelligence³⁶⁷. In Germany, the Federal Ministry of Economic Affairs and Energy has issued a handbook for regulatory sandboxes³⁶⁸.

As to the energy sector, a study on Regulatory Sandboxes conducted in 2023 by the Energy Transition Expertise Centre³⁶⁹ referred to initiatives in Finland, Poland and Slovenia, Germany and Netherlands. Moreover, on 2023 Commission published a staff working document named "Regulatory learning in the EU Guidance on regulatory sandboxes, testbeds, and living labs in the EU, with a focus section on energy" where existing and planned frameworks for the setting up and implementation of regulatory sandboxes in the EU legislation were mentioned³⁷⁰

Earlier to the publication of the AI Act, OECD prepared a report on Regulatory sandboxes in Artificial Intelligence³⁷¹, which refers among other issues to lessons from regulatory sandboxes in fintech including both benefits to regulators, firms and consumers, positive impacts (such as fostering fintech venture capital investment, facilitating market entry, improving communication between regulators and firms, adapting regulatory processes, informing regulation, nurturing internationally harmonized

³⁶³ <https://www.datatilsynet.no/en/regulations-and-tools/sandbox-for-artificial-intelligence/>

³⁶⁴ <https://ico.org.uk/for-organisations/advice-and-services/regulatory-sandbox/>

³⁶⁵ <https://digital-strategy.ec.europa.eu/en/news/first-regulatory-sandbox-artificial-intelligence-presented>

³⁶⁶ <https://www.cnil.fr/fr/bac-sable-donnees-personnelles-la-cnil-lance-un-appel-projets-sur-lintelligence-artificielle-dans>, a list of regulatory sandboxes and relevant initiatives is available at <https://artificialintelligenceact.eu/ai-regulatory-sandbox-approaches-eu-member-state-overview/> (accessed on 17-6-2025)

³⁶⁷ <https://cnpd.public.lu/fr/professionnels/outils-conformite/sandbox.html>

³⁶⁸ https://www.congreso.es/docu/docum/ddocum/dosieres/sleg/legislatura_14/spl_5/pdfs/20.pdf

³⁶⁹ <https://op.europa.eu/en/publication-detail/-/publication/86c18e4c-1ecb-11ee-806b-01aa75ed71a1/language-en>

³⁷⁰ COMMISSION STAFF WORKING DOCUMENT Regulatory learning in the EU Guidance on regulatory sandboxes, testbeds, and living labs in the EU, with a focus section on energy, 2023, https://research-and-innovation.ec.europa.eu/document/download/fc6f35cd-a8d6-4770-ae6e-c09ca85cff8c_en?filename=swd_2023_277_f1.pdf

³⁷¹ OECD (2023), "Regulatory sandboxes in artificial intelligence", *OECD Digital Economy Papers*, No. 356, OECD Publishing, Paris, <https://doi.org/10.1787/8f80a0e6-en>.

sandbox frameworks) and challenges faced (such as inefficient and risky implementation of regulatory sandboxes, inefficient sandbox testing processes, future scalability, regulatory fragmentation).

One of the most important initiatives with regards to EU regulatory sandboxes is the EU Regulatory Sandboxes for AI (EUSAIr)³⁷², a two year project funded by the European Union's Digital Europe program, which will provide support to the implementation of AI regulatory sandboxes across the EU to foster AI innovation and competitiveness, enhance legal certainty for innovators and simplify compliance with the AI Act with the aim to coordinate and support the implementation of AI regulatory sandboxes across the EU. In addition, EU has already established specialized Testing and Experimentation Facilities (TEFs) open to all technology providers across Europe to test and experiment AI solutions both software and hardware products and services³⁷³ with the purpose to support AI developers to place into market Trustworthy AI³⁷⁴. Examples of such TEFs are agrifood TEF, TEF-Health, AI- MATTERS and project "Citcom.AI".

2.5.2.2 Regulations and other frameworks referring to regulatory sandboxes

Regulations and other frameworks referring to regulatory sandboxes are:

AI Act

Regulatory sandboxes are now specifically regulated in articles 57, 58 and 59 of the AI Act and "are being designed as tools for the implementation of AI systems under the EU AI Act"³⁷⁵.

Based on the definition of article 3 (55), AI regulatory sandbox *"means a controlled framework set up by a competent authority which offers providers or prospective providers of AI systems the possibility to develop, train, validate and test, where appropriate in real-world conditions, an innovative AI system, pursuant to a sandbox plan for a limited time under regulatory supervision"*. In accordance with AI Act each member state is required to establish regulatory sandboxes by the 2nd of August 2026.

Article 57 of the AI Act, defines among others the objectives for the development of regulatory sandboxes, their basic parameters (time, coverage etc.), operational conditions, intervention/role of competent authorities (provision of guidance, supervision and support within the AI regulatory sandbox, written proof of activities successfully carried out in the sandbox upon request of the provider). As Ruschemeier mentions³⁷⁶, the AI Act does not lay down legal exceptions or obligations for non-enforcement but defines that subject to specific conditions mentioned therein the authorities will not impose fines for any infringements of the AI Act itself. No exceptions for other areas of law are provided and liability of providers/participants to third parties for any damage caused because of experimentation taking place in the sandbox remains³⁷⁷.

³⁷² <https://eusair-project.eu/>

³⁷³ <https://digital-strategy.ec.europa.eu/en/policies/testing-and-experimentation-facilities>

³⁷⁴ <https://digital-strategy.ec.europa.eu/en/policies/testing-and-experimentation-facilities>

³⁷⁵ M. Boura, The Digital Regulatory Framework through EU AI Act: The Regulatory Sandboxes' Approach, Athens Journal of Law - Volume 10, Issue 3, July 2024 – Pages 385-398,

³⁷⁶ Hannah Ruschemeier, Thinking Outside the Box? Regulatory Sandboxes as a Tool for AI Regulation 2024-04-07 Pre-print, under review, p. 9

³⁷⁷ Hannah Ruschemeier, Thinking Outside the Box? Regulatory Sandboxes as a Tool for AI Regulation 2024-04-07 Pre-print, under review, p. 9

Each member state shall establish a regulatory sandbox either alone or jointly with the competent authorities of other member states. Article 58 of the AI Act lays down that the Commission shall adopt implementing acts specifying the detailed arrangements for the establishment, development, implementation, operation and supervision of the AI regulatory sandboxes, referring to the principles to be included with the purpose to avoid fragmentation and safeguard harmonization of procedures for the development and operation of the sandboxes. As commented by Noberto de Andrade, consistency of rules of governance safeguards open, accessible and equal participation of interested parties while avoiding “sandbox shopping”³⁷⁸.

As such, article 58 deals with the practicalities of regulatory sandboxes’ operation setting the operational prerequisites for their function³⁷⁹. To avoid fragmentation, the Union shall adopt implementing acts specifying detailed arrangements for the establishment, development, implementation, operation and supervision of the AI regulatory sandboxes. The importance of those implementing acts in ensuring consistency and support for the establishment and operation of sandboxes has been highlighted by EDPB and EDPS in their joint opinion 5/2021³⁸⁰.

Article 59 lays down the necessary conditions for further processing of personal data for developing certain AI systems in the public interest in the AI regulatory sandbox among which that of AI system to be developed for safeguarding substantial interest by a public authority or another natural or legal person in energy sustainability. Conditions are analyzed in article 59-1a-j of the AI Act.

As mentioned in article 58 par. 3 of the AI Act “*prospective providers in the AI regulatory sandboxes, in particular SMEs and start-ups shall be directed, where relevant to pre-deployment services such as guidance on the implementation of this Regulation, to other value adding services such as help with standardization documents and certification, testing and experimentation facilities, European Digital Innovation Hubs and centers of excellence*”.

As commented by Arto Lanamäki et al³⁸¹ in their paper based on the insights from the “HI Ethics Forum Webinar: Regulatory Sandboxes under the EU AI Act” despite the structured approach available in AI Act, the exact nature and implementation of regulatory sandboxes remain unclear and issues that should be addressed relate to the definition of the institutional form that they should take, the clarification of role of data and the legal framework of exceptions. As such, significant questions to consider for the establishment of regulatory sandboxes include the definition of the form and capacity of the regulatory sandbox, the organizations’ admittance criteria, the scalability of various regulatory sandboxes, the financial model, the effort to accelerate time to market through participation in regulatory sandboxes, the need for addressing the organisation’s need in the AI supply chain as well

³⁷⁸ The EU Artificial Intelligence (AI) Act, a commentary, Wolters Kluwer, 2025, p. 898

³⁷⁹ The EU Artificial Intelligence (AI) Act, a commentary, Wolters Kluwer, 2025, p. 898

³⁸⁰ EDPB-EDPS Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), 2021, https://www.edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf.

paragraph 63

³⁸¹ Arto Lanamäki, Karin Väyrynen, Fanny Vainionpää, Heidi Hietala, Erkki Tervo, Alex Moltzau, Sophie Weerts, Demian Niemeyer, What to Expect from the Upcoming EU AI Act Sandboxes: Panel Report, Digital Society (2025) 4:42 <https://doi.org/10.1007/s44206-025-00201-x>

as the appropriateness of decisions resulted from regulatory sandboxes in case of sector specific rules³⁸².

OECD Recommendation on Artificial Intelligence

The OECD Recommendation on Artificial Intelligence³⁸³ (hereinafter OECD Recommendation) the first intergovernmental standard on AI, was adopted by the OECD Council meeting at Ministerial level on 22 May 2019 and “*aims to foster innovation and trust in AI by promoting the responsible stewardship of trustworthy AI while ensuring respect for human rights and democratic values*”³⁸⁴.

The following revisions took place in 2023 and 2024. OECD Recommendation focused on policy issues that are specific to AI and aimed to define a standard that can be implemented and disposes enough flexibility to adjust in time and at a fast pace of evolution.

Among the recommendations mentioned which are consistent with the principles also therein referred, OECD Recommendation refers that in order to shape an enabling interoperable governance and policy environment for AI, “*governments should promote an agile policy environment that supports transitioning from the research and development stage to the deployment and operation stage for trustworthy AI systems. To this effect, they should consider using experimentation to provide a controlled environment in which AI systems can be tested, and scaled up, as appropriate [...]*”.

Council of Europe Framework on Artificial Intelligence and Human Rights, Democracy and the Rule of Law³⁸⁵

In addition to abovementioned regarding the importance of the Convention, in article 13 the importance of safe innovation is highlighted through the establishment of experimenting and testing artificial intelligence systems under the supervision of each party to the convention competent authorities.

Commission Recommendation on renewable energy projects and PPCs

The Commission Recommendation on speeding up permit-granting procedures for renewable energy projects and facilitating Power Purchase Agreements_C (2022)3219³⁸⁶ in 2022, noted that sandboxes could support innovative decarbonization technologies needed for climate neutrality. As mentioned, setting up regulatory sandboxes, the testing, in a real-life environment, of innovative technologies, products, services or approaches, which are not fully compliant with the existing legal and regulatory framework, could support innovation and facilitate the subsequent adaptation of the regulatory environment to accommodate them³⁸⁷.

³⁸² Arto Lanamäki, Karin Väyrynen, Fanny Vainionpää, Heidi Hietala, Erkki Tervo, Alex Moltzau, Sophie Weerts, Demian Niemeyer, What to Expect from the Upcoming EU AI Act Sandboxes: Panel Report, Digital Society (2025) 4:42 <https://doi.org/10.1007/s44206-025-00201-x>

³⁸³ OECD Recommendation on Artificial Intelligence, available at <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449>

³⁸⁴ OECD Recommendation on Artificial Intelligence, available at <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449>

³⁸⁵ <https://rm.coe.int/1680afae3c>

³⁸⁶ [https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=intcom:C\(2022\)3219](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=intcom:C(2022)3219)

³⁸⁷ Paragraph 18

As such, through the Recommendation, member states were encouraged to establish regulatory sandboxes to grant targeted exemptions from national, regional or local legislative or regulatory framework for innovative technologies, products, services or approaches, to facilitate permit-granting in support of the deployment and system integration of renewable energy, storage, and other decarbonization technologies in line with Union legislation³⁸⁸.

Directive (EU) 2023/2413 for the promotion of energy from renewable energy sources

According to Directive 2023/2413³⁸⁹ "*Member States shall promote the testing of innovative renewable energy technology for producing, sharing and storing of renewable energy through pilot projects in a real-world environment, for a limited period, in accordance with the applicable Union law and accompanied by appropriate safeguards to ensure the secure operation of the energy system and avoid disproportionate impact on the functioning of the internal market, under the supervision of a competent authority*".

2.5.2.3 Testing of high risks AI systems in real world conditions outside AI regulatory sandboxes

For the providers or prospective providers of high-risk AI systems to speed the process of development and placing on the market of their systems, they should be able to take advantage of testing on real world conditions, without participating in an AI regulatory sandbox³⁹⁰.

The AI Act also includes a mechanism for testing AI systems in real – world conditions. According to AI Act's definitions, testing in real world conditions means "*the temporary testing of an AI system for its intended purpose in real-world conditions outside a laboratory or otherwise simulated environment, with a view to gathering reliable and robust data and to assessing and verifying the conformity of the AI system with the requirements of this Regulation and it does not qualify as placing the AI system on the market or putting it into service within the meaning of this Regulation, provided that all the conditions laid down in Article 57 or 60 of AI Act are fulfilled*"³⁹¹.

Real-world testing may take place both in and outside regulatory sandboxes and article 60 of the AI Act applies in cases of real-world testing outside the regulatory sandbox³⁹² (entering into force on 2 August 2026). Literature has commented that testing under real conditions outside regulatory sandboxes weakens the potential of regulatory learning since the latter is not explicitly mentioned as an objective³⁹³.

Real-world testing outside the regulatory sandbox may be conducted by providers or prospective providers of high risk AI systems listed in Annex III of the AI Act, permitted only in accordance with article 60 and the real world testing plan as defined in article 3 (53) of the AI Act and whose elements will be specified by implementing acts of Commission.

Requirements set by article 60, refer among others to the appropriate time of real world testing before the placing on the market or putting into service, the prerequisite of ethical review when applicable,

³⁸⁸ Paragraph 32

³⁸⁹ https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202302413

³⁹⁰ Recital 141 of the AI Act

³⁹¹ Article 3 par. 57 of AI Act

³⁹² The EU Artificial Intelligence (AI) Act, a commentary, Wolters Kluwer, 2025, p. 929

³⁹³ Hannah Ruschemeier, Thinking Outside the Box? Regulatory Sandboxes as a Tool for AI Regulation, 2024-04-07 Pre-print, under review, p. 10

the development of a real world testing plan and its submission to the market surveillance authority, the approval of the testing and the plan by the market surveillance authority or its tacit approval after lack of answer within 30 days, the registration of the testing at national level, the establishment in Union other appointment of a legal representative, the adoption of appropriate guarantees in case of transfer of processed data to third countries, the limited duration of the testing (less than 6 months), protection of participating subjects that are vulnerable, information of the deployer or prospective deployer by the provider and the drafting of an agreement between them, the informed consent of the subject regarding its participation or specific conditions when the seeking of consent would prevent the AI system from being tested, oversight of the testing by the necessary persons, the possibility that predictions and recommendations or decisions of the AI system can be reversed and disregarded.

Finally, according to article 61 of the AI Act, taking in account the possible consequences of such testing to individuals, prior freely given informed consent should be obtained from data subject prior to their participations in such testing on condition that the information specified in the article have been provided.

Market surveillance authorities are given the authority to supervise the tests and stop them when needed. However, unlike the sandbox procedure, real world tests outside a sandbox are not necessarily integrated with data protection enforcement and data protection requirements can be enforced normally, without the restrictions set by sandboxes³⁹⁴.

2.5.2.4 Union AI testing support structures

According to article 84 of the AI Act, the Commission shall appoint one or more Union AI testing support structures to carry out the tasks listed under article 21 (6) of Regulation 2019/1020 regarding carrying out testing of products, provide independent technical and scientific advice and develop new techniques and methods of analysis in the area of AI.

As mentioned in article 84 “[...] *Union AI testing support structures shall also provide independent technical or scientific advice at the request of the Board, the Commission, or of the market surveillance authorities*”. As mentioned by Zegler³⁹⁵, this article indicates that the Union’s AI aim regarding testing support structures is to provide space where within a testing environment AI systems can be tested, new techniques and methods of analysis in AI area and methods of analysis can be developed.

2.5.2.5 Difficulties of regulatory sandboxes in practice

However, several obstacles may hinder the creation of a regulatory sandbox or the participation of stakeholders and may cause it to deviate from its original objectives. The most notable challenges include:

- Absence of a specific, unified legal framework
- Requested legal deviations falling outside the jurisdiction of the implementing authority
- Participation limitations, especially for smaller companies
- Lack of resources and expertise from the authorities attempting to establish a sandbox³⁹⁶

³⁹⁴ Marco Almada, Law and Compliance in AI Security & Data Protection, AI and Data Protection Module, p. 214

³⁹⁵ The EU Artificial Intelligence (AI) Act, a Commentary, Wolters Kluwer, p. 1186

³⁹⁶ Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation, Natalia Díaz-Rodríguez et al., Information Fusion (2023)

These difficulties may lead to negative consequences. A frequently cited drawback is market disruption, as sandboxes allow some participants to deviate from regulations while others cannot, potentially leading to long-term competitive imbalances. Consumers may also be negatively affected, especially if safety concerns arise, such as the mishandling of personal data.

On the positive side, sandboxes can:

- Promote business innovation
- Encourage innovation in regulation by identifying needs for regulatory updates or new frameworks
- Enhance cooperation and knowledge exchange among regulators, administrative authorities, and stakeholders
- Provide useful insights for revising regulations

In conclusion, regulatory sandboxes are a vital tool for promoting innovation while ensuring compliance with legislation. When properly implemented, they can deliver substantial benefits to both businesses and consumers.

2.5.3 Ethical Considerations and Societal Impact of AI

As artificial intelligence (AI) becomes increasingly embedded in our lives, affecting directly or indirectly many of our daily activities across diverse domains—from entertainment and education to justice and healthcare—the issues arising from its use highlight the need to examine its ethical dimensions and evaluate its risks, in order to ensure the protection of human rights.

AI ethics, a sub-field of applied ethics, is “*a set of values, principles and techniques that employ widely accepted standards of right and wrong to guide moral conduct in the development and use of AI Technologies*”³⁹⁷.

AI ethics are necessary since misuse, abuse, bad design or negative unintended consequences of AI systems can lead to important negative impacts to persons and society³⁹⁸.

Moreover, trustworthiness is a prerequisite for people and societies to develop, deploy, and use AI systems. *Without AI systems – and the human beings behind them – being demonstrably worthy of trust, unwanted consequences may ensue, and their uptake might be hindered, preventing the realization of the potentially vast social and economic benefits that they can bring*³⁹⁹.

Some of the ethical challenges of AI include, among other issues regarding transparency and black box effects, bias/discriminatory effects, autonomy issues, security and resilience, and ethical control⁴⁰⁰. More specifically, AI systems have the capacity to reproduce and increase the patterns of discriminatory

³⁹⁷ Leslie, D. (2019). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute.

<https://doi.org/10.5281/zenodo.3240529>

³⁹⁸ Matthew G. Hanna et al, Ethical and Bias Considerations in Artificial Intelligence/Machine Learning, Mod Pathol (38 (2025), <https://doi.org/10.1016/j.modpat.2024.100686>

³⁹⁹ HLEG guidelines, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

⁴⁰⁰ Dr Mihalis Kritikos, Ethical aspects of AI, the viewpoint of the EU, 14-3-2024 ppt presentation <https://talos-ai4ssh.uoc.gr/wp-content/uploads/2019/04/Kritikos-SLIDES.pdf>

phenomena that appear in the society in which they are created and used⁴⁰¹ while at the same time a risk of discrimination exists when algorithms are used for purposes such as to profile persons or to resolve situations in criminal justice⁴⁰².

In addition, AI raises privacy considerations due to capacity to access huge volume of data about persons and process it with quickly, enabling forecasts about a person's behavior, state of mind and identity by sensing information that is not necessarily defined as personal data resulting in invasion to privacy⁴⁰³.

AI will augment the number of predictions that can take place thus eroding humans' s ability to give a causal interpretation to outcomes, in such a way that the notions of transparency, human control, accountability and liability over results will be highly questioned⁴⁰⁴ . Moreover, inserting AI into the system raises concern, including issues related to algorithmic bias, workforce displacement, and inequities in society members⁴⁰⁵. As mentioned in Artificial intelligence ante portas: Legal and ethical reflections, *"AI's potential for empowerment, the risks associated with entering into "filter bubbles" or with using methodologies for social scoring using AI and the affordability and accessibility of AI services are associated with concerns about human safety, health and security, freedom, privacy, integrity, dignity, self-determination and non – discrimination"*⁴⁰⁶.

AI systems are increasingly used to assist in important decision-making processes, such as selecting the most suitable resumes for job applications, diagnosing illnesses through imaging, or determining whether a bank should grant a loan⁴⁰⁷. As a result, concerns arise regarding whether an AI system can be truly impartial and fair, since its decisions depend on the data it has been trained on, which are

⁴⁰¹ Leslie, D., Burr, C., Aitken, M., Cowls, J., Katell, M., and Briggs, M. (2021). Artificial intelligence, human rights, democracy, and the rule of law: a primer. The Council of Europe.

⁴⁰² EPRS, European Parliamentary Research Service, 2019, EU guidelines on ethics in artificial intelligence: context and implementation
[https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI\(2019\)640163_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI(2019)640163_EN.pdf)

⁴⁰³ Leslie, D., Burr, C., Aitken, M., Cowls, J., Katell, M., and Briggs, M. (2021). Artificial intelligence, human rights, democracy, and the rule of law: a primer. The Council of Europe.

⁴⁰⁴ EDPB-EDPS Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act),
https://www.edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf, paragraph 3

⁴⁰⁵ Md Morshed Alam et al, Artificial intelligence integrated grid systems: Technologies, potential frameworks, challenges, and research directions <https://doi.org/10.1016/j.jser.2024.115251>

⁴⁰⁶ Mihalīs Kritikos, Scientific Foresight Unit (STOA), EPRS, European Parliamentary Research Service, Artificial Intelligence *ante portas*: Legal & ethical reflections,
[https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/634427/EPRS_BRI\(2019\)634427_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/634427/EPRS_BRI(2019)634427_EN.pdf)

⁴⁰⁷ Understanding Relations Between Perception of Fairness and Trust in Algorithmic Decision Making, Jianlong Zhou et al., The International Conference on Behavioral and Social Computing (BESC 2021)

often biased and reproduce existing social inequalities⁴⁰⁸. One such example of racial discrimination is the use of the CalGang database, which is biased against people of color, to train a model aimed at predicting criminal activity⁴⁰⁹.

For AI systems to function effectively, the collection and processing of large volumes of personal data is required, not only during initial training. This underscores the importance of protecting personal data and managing it properly against malicious use⁴¹⁰. It also leads to a lack of trust from users and concerns about the safeguarding of their privacy, as seen, for example, in AI assistants that collect data such as location history or purchase records⁴¹¹.

Furthermore, there are cases where the main function of AI systems infringes on fundamental human rights, aiming instead at oppression and control. We have seen AI used in border control with facial recognition or biometric identification applications⁴¹², or to target specific social minorities, such as the Uyghurs in China, who are subject to continuous surveillance and personal data collection for malicious purposes⁴¹³. This category also includes military AI applications, which eliminate human oversight in decisions involving matters of life and death.

A particularly significant issue of concern, which is not discussed proportionally to its seriousness, is the environmental impact of AI systems. Although AI can actively contribute to finding solutions to climate change through improved climate forecasting, optimized resource management, and broader support for related research, the training of machine learning algorithms requires substantial computational resources, which translates into increased energy consumption and carbon dioxide emissions⁴¹⁴.

According to literature, AI technologies in the energy sector have raised concerns among other in safety, privacy, transparency, fairness and accountability⁴¹⁵. The hunger of many AI applications for data may lead to privacy infringements when increasingly more fine-grained forecasts and predictions of consumer energy demands are made possible by these technologies, while at the same time to be able to provide accurate predictions for demand may require various kinds of personal data or data that

⁴⁰⁸ The Challenges of Algorithm-Based HR Decision-Making for Personal Integrity, Ulrich Leicht-Deobal, *Journal of Business Ethics* (2019) 160:377–392

⁴⁰⁹ Towards Risk-Free Trustworthy Artificial Intelligence: Significance and Requirements, *International Journal of Intelligent Systems*, 2023, 4459198

⁴¹⁰ Towards Risk-Free Trustworthy Artificial Intelligence: Significance and Requirements, *International Journal of Intelligent Systems*, 2023, 4459198

⁴¹¹ Okay google, what about my privacy?: User's privacy perceptions and acceptance of voice based digital assistants, M. Vimalkumar, *Computers in Human Behavior* (2021)

⁴¹² Unpacking AI-enabled border management technologies in Greece: To what extent their development and deployment are transparent and respect data protection rules?, Eleftherios Chelioudakis, *Computer Law & Security Review* (2024)

⁴¹³ Artificial intelligence (AI) and human rights: Using AI as a weapon of repression and its impact on human rights, H. Akin ÜNVER (2024)

⁴¹⁴ Survey on AI Ethics: A Socio-technical Perspective, Dave Mbiazi et al, (2023)

⁴¹⁵ Noorman, M.; EspinosaApráez, B.; Lavrijssen, S. AI and Energy Justice. *Energies* 2023, 16, 2110.

<https://doi.org/10.3390/en16052110> (cross reference to) Niet, I.; van Est, R.; Veraart, F. Governing AI in Electricity Systems: Reflections on the EU Artificial Intelligence Bill. *Front. Artif. Intell.* 2021, 4, 690237.

can reveal a lot about a person's daily schedule and habits.⁴¹⁶ Transparency may also be challenged when the complexity of energy systems increases because of the integration of AI technologies. The black box character of some AI technologies can be a contributing factor, while energy systems may be less transparent when the AI development is assigned to private companies and start-ups also makes energy systems less transparent. *"Moreover, AI-based energy systems that perform a large part of their tasks autonomously can make it more difficult for human operators to intervene effectively because of their complexity and the speed with which they perform their operations. When price negotiations are decentralized—supported by autonomously acting AI-based systems—competition authorities and energy regulators will have a hard time monitoring how prices develop. Moreover, they will have limited means and powers to control these developments"*⁴¹⁷

Recent AI boom resulted in the development of a whole body of ethical guidelines to be adopted to the extent possible by technology developers⁴¹⁸ giving rise to "AI Ethics boom" i.e. a period of increased demand for regulation and normative guidance in the field⁴¹⁹. Given the impact of AI on society, many organizations have presented a variety of initiatives to establish ethical principles for the embracing of socially beneficial AI⁴²⁰.

Two public repositories regarding AI principles are available: a) the AI Ethics Guidelines Global Inventory, a project by AlgorithmWatch that mapped frameworks that seek to set out principles of how systems for automated decision-making (ADM) can be developed and implemented ethically (last update in April 2020), including 167 guidelines, available online as archive⁴²¹ and Linking Artificial Intelligence Principles, an initiative and platform for integrating, analyzing, and promoting global Artificial Intelligence Principles and their social and technical practices World Wide, from different stakeholders such as research institutes, non-profit organizations, non-governmental organizations, companies, etc.⁴²².

Some of the AI ethics guidelines (the list is not restrictive) are the European Commission's High level-Expert Group on Artificial Intelligence (HLEG Guidelines), Report on the Future of Artificial Intelligence,

⁴¹⁶ Noorman, M.; Espinosa

Apráez, B.; Lavrijssen, S. AI and Energy Justice. *Energies* 2023, 16, 2110. <https://doi.org/10.3390/en16052110>

⁴¹⁷ Noorman, M.; Espinosa

Apráez, B.; Lavrijssen, S. AI and Energy Justice. *Energies* 2023, 16, 2110. <https://doi.org/10.3390/en16052110>

⁴¹⁸ Thilo Hagendorff, The Ethics of AI Ethics: An Evaluation of Guidelines, *Minds and Machines* (2020) 30:99–120 <https://doi.org/10.1007/s11023-020-09517-8>

⁴¹⁹ Nicholas Kluge Correia, Camila Galvão, James William Santos, ..., Luiza Galvão, Edmund Terem, Nythamar de Oliveira, Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance, <https://pdf.sciencedirectassets.com/776857/1-s2.0-S2666389922X00111/1-s2.0-S2666389923002416/main.pdf>

⁴²⁰ Luciano Floridi and Josh Cowls, A Unified framework of five Principles for AI in Society, 2 June 2019, <https://hdr.mitpress.mit.edu/pub/l0jsh9d1/release/8>

⁴²¹ <https://inventory.algorithmwatch.org/>

⁴²² <https://www.linking-ai-principles.org/>

Beijing AI Principles, OECD Recommendation of the Council of Artificial Intelligence, the malicious use of artificial intelligence, AI4People, The Asilomar AI Principles, AI Now 2016 Report, AI Now 2017 Report, AI Now 2018 Report, AI Now 2019 Report, Principles for Accountable Algorithms and a Social Impact Statement for Algorithms, Montreal Declaration for Responsible Development of Artificial Intelligence, Open AI Charter, Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems (First Edition), ITI AI Policy Principles, Microsoft AI principles, DeepMind Ethics & Society Principles, Artificial Intelligence at Google, Everyday Ethics for Artificial Intelligence, Partnership on AI⁴²³.

The increase in Guidelines drafting regarding AI has been to such an extent that specific literature has been published to gather, compile, and compare them being a meta-analysis of published AI Ethical guidelines⁴²⁴. In the article published in 2019 by L. Floridi et al, named "*A unified framework of Five principles for AI in Society*"⁴²⁵, five core principles for ethical AI were identified through comparative analysis of six high -profile initiatives established in the interest of socially beneficial AI and more specifically the Asilomar AI principles developed on January 2017, the Montreal Declaration for Responsible AI⁴²⁶, the Ethically Aligned Design: A vision for Prioritizing Human Wellbeing with Autonomous and Intelligent Systems⁴²⁷, the Statement on Artificial Intelligence, Robotics and "Autonomous" Systems by the European Group on Ethics in Science and New Technologies published on March 2018⁴²⁸, the "AI in the UK: ready, willing and able?" drafted by House of Lords, Select Committee on Artificial Intelligence and published in⁴²⁹, the Tenets of Partnership on⁴³⁰. These overarching principles which emerged from the analysis are beneficence, non-maleficence, autonomy, justice, and explicability.

In the article of N. Klunge Correa et al, "Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance"⁴³¹, the distribution of principles among the 200 documents examined, indicated that the most recurrent five principles were: transparency, explainability/

⁴²³ A full table with key issues and principles included can be found at Thilo Hagendorff, The Ethics of AI Ethics: An Evaluation of Guidelines, Minds and Machines (2020), Springer, 30:99–120 <https://doi.org/10.1007/s11023-020-09517-8>

⁴²⁴ Thilo Hagendorff, The Ethics of AI Ethics: An Evaluation of Guidelines, Minds and Machines (2020), Springer, 30:99–120 <https://doi.org/10.1007/s11023-020-09517-8>, Luciano Floridi and Josh COWLS, A Unified framework of five Principles for AI in Society, 2 June 2019, <https://hdr.mitpress.mit.edu/pub/l0jsh9d1/release/8>, Nicholas Kluge Correˆa, Camila Galvaˆo, James William Santos, ..., Luiza Galvaˆo, Edmund Terem, Nythamar de Oliveira, Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance, Jobin, A., Ienca, M., and Vayena, E. (2019). The global landscape of AI ethics guidelines. Nat. Mach. Intell. 1, 389–399

⁴²⁵ <https://hdr.mitpress.mit.edu/pub/l0jsh9d1/release/8>

⁴²⁶ https://declarationmontreal-iaresponsable.com/wp-content/uploads/2023/04/UdeM_Decl-IA-Resp_LA-Declaration-ENG_WEB_09-07-19.pdf

⁴²⁷ https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf

⁴²⁸ <https://op.europa.eu/en/publication-detail/-/publication/dfebe62e-4ce9-11e8-be1d-01aa75ed71a1#>

⁴²⁹ <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf>

⁴³⁰ <https://partnershiponai.org/about/#tenets>

⁴³¹ Patterns 4, 100857, October 13, 2023,

auditability, reliability/ safety, security/ trustworthiness, justice/equity/fairness/non-discrimination, privacy, accountability/liability⁴³².

Although as mentioned above the list of guidelines and principles for ethical AI is very extensive, for the purposes of this project, the following will be analyzed because in our view their importance is highlighted mainly through the AI Act.

2.5.3.1 *Ethics Guidelines for Trustworthy AI, High-Level Expert Group on Artificial Intelligence*

As specifically mentioned in the AI Act, without prejudice to the legally binding requirements of AI Act and any other applicable Union law, Ethical Guidelines for trustworthy AI (hereinafter "HLEG guidelines"⁴³³) contribute to the design of coherent, trustworthy and human-centric AI, in line with the Charter and with the values on which the European Union is founded⁴³⁴ while *"providers and, as appropriate, deployers of all AI systems, high risk or not and AI models should also be encouraged to apply on a voluntary basis additional requirements related, for example the elements of the Union's Ethics Guidelines for Trustworthy AI[...]"*⁴³⁵.

HLEG Guidelines, especially through the practical assessment list included in Chapter III, are tools that can be used voluntarily by AI stakeholders to achieve trustworthy AI when developing, deploying, or using AI systems in the EU⁴³⁶. HLEG Guidelines, having built on the work of the European Group on Ethics in Science and New Technologies which proposed 9 principles based on the fundamental values of EU Treaties and the Charter, set out the framework for achieving trustworthy AI by analyzing the two out of three components of trustworthy AI meaning the component of ethical (meaning adherence to ethical principles and values) and robust (from a technical and social perspective to avoid harm), leaving outside of the scope the component of lawfulness (meaning compliance with applicable laws and regulations), which should all be met throughout the system's entire life cycle with the purpose to foster responsible and sustainable AI innovation in Europe⁴³⁷.

According to HLEG Guidelines, the four ethical principles that must be respected to ensure that AI systems are developed and used in a trustworthy manner are the following⁴³⁸:

- a) Respect for human autonomy, meaning that humans interacting with AI systems must be able to maintain self-determination over themselves,
- b) Prevention of harm, meaning that AI systems should not harm or adversely affect human beings and should safeguard human dignity including mental and physical integrity.
- c) Fairness, meaning among others that AI systems should be committed to ensuring equal and just distribution of benefits and cost and should also ensure that individuals and groups should not be subject to unfair bias and discrimination, and

⁴³² Patterns 4, 100857, October 13, 2023 , p. 11

⁴³³ <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

⁴³⁴ Recital 27 of the AI Act

⁴³⁵ Recital 165 of AI Act

⁴³⁶ EPRS, European Parliamentary Research Service, 2019, EU guidelines on ethics in artificial intelligence: context and implementation

[https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI\(2019\)640163_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI(2019)640163_EN.pdf)

⁴³⁷ Ethical Guidelines for trustworthy AI, High-Level Expert Group on Artificial Intelligence, p.5

⁴³⁸ Ethical Guidelines for trustworthy AI, High-Level Expert Group on Artificial Intelligence, p. 12

- d) Explicability, meaning that processes should be transparent, AI systems' capacities and objectives should be openly communicated, and decisions should be explainable to the extent possible to person directly and indirectly affected avoiding black box cases.

For the implementation of trustworthy AI, a non-exhaustive list of seven requirements built on the abovementioned four principles is laid down and has as follows⁴³⁹:

1. Human agency and oversight, including protection of fundamental rights. Users should be given the knowledge to comprehend and interact with AI systems, avoiding being subject to a decision based on automated processing when this significantly affects them. In addition, for human autonomy to be respected, there should be mechanisms that ensure human oversight. The rationale behind this was that the allocation of functions between humans and AI systems should follow principles as based on human centism, enabling human choice opportunities and stressing the importance of human oversight over AI systems work processes⁴⁴⁰
2. Technical robustness and safety including resilience to attack and security fall black plan and general safety, accuracy, reliability and reproducibility, meaning that AI systems should be developed to prevent risks and reliably behave as intended while minimizing unintentional and unexpected harm and be protected against vulnerabilities that may enable them to be exploited by adverse events. As such they should be resilient to attack and security, have safeguards to make use of a fallback plan when needed, and produce accurate judgments and predictions, while their results should be accurate and reliable.
3. Privacy and data governance including respect for privacy, quality and integrity of data and access to data, meaning that AI systems should secure privacy and data protection throughout a system's entire lifecycle, including data provided by the user and generated about the user. At the same time, the quality of data should guarantee avoidance of biases, inaccuracies and mistakes, and integrity of data should also be ensured avoiding malicious data to be inserted in the system and processes and data sets used must be tested and documented at each step (planning, training, testing and deployment). Finally, data governing protocols should be established to secure appropriate and authorized access to the data.
4. Transparency including traceability, explainability and communication, meaning that data sets, processes, data gathering, data labelling and algorithms used by the AI system as well as decisions made by it, should be documented to allow traceability and transparency. At the same time the ability to explain technical processes of an AI system and human related decisions should also exist. In addition, AI systems must be identifiable as such and not represent themselves as humans.
5. Diversity, non-discrimination and fairness including the avoidance of unfair bias, accessibility and universal design and stakeholder participation, meaning that identifiable and discriminatory bias should be removed in the collection phase of data when possible and take measures to avoid bias in the way the system is developed. Also, systems should enable equitable access

⁴³⁹ Ethical Guidelines for trustworthy AI, High-Level Expert Group on Artificial Intelligence, p. 14

⁴⁴⁰ 'Human oversight' in the EU artificial intelligence act: what, when and by whom? Lena Enqvist, LAW, INNOVATION AND TECHNOLOGY 2023, VOL. 15, NO. 2, 508 -535
<https://doi.org/10.1080/17579961.2023.2245683>, p. 512

and participation of all people in existing and emerging computer-mediated human activities. Finally, consultation with stakeholders probably affected is advisable.

6. Societal and environmental well-being, including sustainability and environmental friendliness, social impact, society and democracy meaning that AI systems should also be sustainable and environmentally friendly, while societal impact of the systems should also be monitored and considered. Assessment of a system should also consider its impact with regards to democracy and society at large.
7. Accountability including auditability, minimization and reporting of negative impact, tradeoffs and redress necessitates that mechanisms to ensure responsibility and accountability of AI systems are in place at the development, deployment and use. Assessment of algorithms, data and design processes (auditability), identification, documentation, assessment and minimization of potential negative impacts should be in place, and redress should be possible when adverse impact occurs. In addition, reports on actions that lead to specific system outcomes and response to consequences of such outcomes should be safeguarded, while trade-offs while implementing the requirements should be rationally addressed.

The above mentioned seven requirements can be implemented through technical and non-technical methods engaged throughout the whole life cycle of an AI system⁴⁴¹. The AI HLEG developed seven non-binding ethical principles for AI which are intended to help ensure that AI is trustworthy and ethically sound⁴⁴². The last chapter of Guidelines lays down a non-exhaustive Trustworthy AI assessment list to operationalize Trustworthy AI that was firstly presented in its piloting phase closed on 1 December 2019, and then after feedback was received, it was then revised and presented as the final Assessment List for Trustworthy AI in July 2020. This list aims to help assess whether the AI system that is being developed, deployed, produced, or used adheres to seven abovementioned requirements for trustworthy Artificial Intelligence.

2.5.3.2 *Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment*

The assessment List for Trustworthy Artificial Intelligence (hereinafter "ALTAI") for self-assessment is a tool developed for the support of AI developers and deployers to achieve Trustworthy AI, updated after the revision of its piloting process⁴⁴³.

The list includes AI principles mentioned in HLEG Guidelines, incorporated in a list of questions and is available in pdf and as a web-based tool. ALTAI is intended for flexible use, and elements can be added according to the sector in question. As mentioned in ALTAI, prior to self-assessment an AI system with the list, a fundamental rights impact assessment should be performed (FRIA) with questions drawing on specific articles in the Charter and the European Convention on Human Rights (ECHR), its protocols and the European Social Charter⁴⁴⁴.

Based on the seven requirements analyzed in the Guidelines, numerous questions are listed about each of the abovementioned requirements.

⁴⁴¹ Ethical Guidelines for trustworthy AI, High-Level Expert Group on Artificial Intelligence, p. 20

⁴⁴² Recital 27 of the AI Act

⁴⁴³ <https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment>

⁴⁴⁴ The assessment list for trustworthy Artificial Intelligence (ALTAI) for self assessment, p. 5

As such, the subsection on human agency and autonomy deals with the effect of AI systems whose purpose is to influence or support humans in the process of decision taking⁴⁴⁵.

In addition, it relates to the outcome of humans when dealing with AI systems that behave as humans and AI systems' outcome on human affection and confidence. The subsection of human oversight deals with mechanisms adopted to secure human oversight (human in the loop, human in the loop, and human in command).

The requirement of technical robustness and safety is assessed through questions that address four main issues namely security, safety, accuracy and reliability, fall back plans and reproducibility⁴⁴⁶. The requirement of Privacy and data governance is assessed through questions related to the mechanism used to ensure quality and integrity of data used, appropriateness and minimization of data in relation to the domain of AI system and processing of data in accordance with data protection obligations⁴⁴⁷.

The requirement of transparency is assessed through examination of traceability of both data used and AI models or rules, explainability meaning the ability to explain technical processes of the AI system and the reasoning behind the decisions or predictions, and open communication about the capabilities and limitations of the AI system⁴⁴⁸ including system's level of accuracy and/or error rates.

The requirement of diversity, non-discrimination and fairness is assessed through questions related to measures taken to avoid unfair bias regarding both the use of input data and algorithm design, diversity and representativeness of end users in the data and mechanisms in place to enable flagging of issues related to bias⁴⁴⁹.

Accessibility of the systems to all people and consideration of Universal Design principles and stakeholder participation are also examined. The requirements of societal and environmental well-being are also examined through assessment of the potential negative impacts of the AI system on the environment and measures to reduce environmental impact.

In addition, the impact of the AI system in a working environment is also assessed in parallel with its impact from a societal perspective. Finally, the requirement of accountability is also assessed through questions related to mechanisms in place to ensure responsibility for the development and use of AI systems through the ability of the system to be independently auditable⁴⁵⁰. Finally risk management mechanisms through the possibility to identify, assess, document and minimize the potential negative impact of AI systems are also examined.

⁴⁴⁵ The assessment list for trustworthy Artificial Intelligence (ALTAI) for self assessment, p. 8

⁴⁴⁶ The assessment list for trustworthy Artificial Intelligence (ALTAI) for self assessment, p. 9

⁴⁴⁷ The assessment list for trustworthy Artificial Intelligence (ALTAI) for self assessment, p. 12

⁴⁴⁸ The assessment list for trustworthy Artificial Intelligence (ALTAI) for self assessment, p. 14

⁴⁴⁹ The assessment list for trustworthy Artificial Intelligence (ALTAI) for self assessment, p. 16

⁴⁵⁰ The assessment list for trustworthy Artificial Intelligence (ALTAI) for self assessment, p. 21

2.5.3.3 AI Act

Literature has commented that HLEG Guidelines have influenced AI Act by inspiring the obligations for high-risk AI systems and other risk categories, functioning as normative guidance for the development of voluntary codes of conduct and “general normative principles” for all AI practitioners⁴⁵¹.

More specifically, as Smuha commented, article 27 lays down the cases where a fundamental rights impact assessment is necessary for high risk AI systems, article 14 has been commented as transposing the requirement of human oversight and requirements to secure it are listed, article 15 has been commented as introducing the requirement of technical robustness and safety through achievement of appropriate level of accuracy, robustness and cybersecurity in the design and development of high risk AI systems, article 10 has been commented as incorporating the requirement of privacy and data governance, the requirement of transparency has been commented to be incorporated in articles 11 and 12 regarding technical documentation and record keeping, article 13 regarding transparency and provision of information to deployers, article 10 and 16 partially incorporate diversity, non-discrimination and fairness, while societal and environmental well-being were considered to be partially inserted in article 5 regarding prohibition of subliminal manipulation.⁴⁵² Finally, article 9 regarding the establishment of a risk management system in relation to high-risk AI systems may be considered as implementation of HLEG Guidelines’ principle of accountability⁴⁵³. HLEG’s Guidelines influence in AI Act is also evident through the adoption of article 95 referring to voluntary codes of conduct for systems other than high risk, where one of the elements to be included is applicable elements provided for in Union ethical guidelines for trustworthy AI, presumably referring to HLEG Guidelines⁴⁵⁴.

2.5.3.4 Recommendation on the Ethics of Artificial Intelligence, UNESCO

The Recommendation on the Ethics of Artificial Intelligence⁴⁵⁵, adopted on 23 November 2021, is addressed to Member states, AI actors and authorities responsible for creating legal and regulatory frameworks throughout the complete AI system life cycle, and can be used as ethical guidance to both public and private sectors, by providing a basis for the conduct of ethical impact assessment of the AI systems⁴⁵⁶.

The values of respect, protection and promotion of human rights and fundamental freedoms, and human dignity are essential throughout the life cycle of AI systems and must be respected and protected. Environmental and ecosystem flourishing should be encouraged. AI Systems should respect and promote diversity and inclusiveness while at the same time AI systems should contribute to the interconnectedness of all living creatures with each other and to peaceful living while inclusiveness, justice and equity should also be promoted.

⁴⁵¹ Smuha, Nathalie A., The Work of the High-Level Expert Group on AI as the Precursor of the AI Act (October 01, 2024). Available at SSRN: <https://ssrn.com/abstract=5012626> or <http://dx.doi.org/10.2139/ssrn.5012626> p. 10

⁴⁵² Smuha, Nathalie A., The Work of the High-Level Expert Group on AI as the Precursor of the AI Act (October 01, 2024). Available at SSRN: <https://ssrn.com/abstract=5012626> or <http://dx.doi.org/10.2139/ssrn.5012626>

⁴⁵³ Nathalie A. Smuha, The Work of the High-Level Expert Group on AI as the Precursor of the AI Act, p. 11-15

⁴⁵⁴ Smuha, Nathalie A., The Work of the High-Level Expert Group on AI as the Precursor of the AI Act (October 01, 2024). Available at SSRN: <https://ssrn.com/abstract=5012626> or <http://dx.doi.org/10.2139/ssrn.5012626> p. 16

⁴⁵⁵ <https://unesdoc.unesco.org/ark:/48223/pf0000381137>, adopted on 23 November 2021

⁴⁵⁶ <https://unesdoc.unesco.org/ark:/48223/pf0000381137>, adopted on 23 November 2021, p. 11

In addition, several principles that should rule AI systems are mentioned among which the proportionality and do no harm relevant also to the choice to use AI systems and AI method to be used, safety and security referring to the avoidance, prevention and elimination of safety and security risks, fairness and non-discrimination with regards to ensuring that AI technologies' advantages are openly available and accessible, taking under consideration also different needs of different groups and people while fairness should be ensured by minimizing discriminatory or biased applications and outcomes.

Moreover, sustainability of societies should be promoted. The right to privacy and compliance with data protection frameworks and governance mechanisms should be established so that AI systems ensure that personal information is protected throughout the life cycle of an AI system. Human oversight and determination should be available so that it can be possible to attribute ethical and legal responsibility at any stage of the AI system life cycle. AI systems should also be subject to transparency and explainability appropriate to the context and impact, with the former enabling people to understand how each stage of the AI system is put in place, while with the latter providing insights among others into the outcome of AI systems.

Responsibility and accountability should also be ensured for decisions and actions based in AI systems and should be attributable to AI actors according to their role. Finally, public awareness and literacy should also be promoted, so that all members of society are appropriately informed to take decisions about the use of AI systems, while multi-stakeholder and adaptive governance and collaboration should be promoted for the benefits of AI to be shared by all.

2.5.3.5 Recommendation of the OECD on Artificial Intelligence

In addition to what above mentioned about the Recommendation on Artificial Intelligence, the Council recommends that members and non-members adhering to the Recommendation promote and implement the principles included to ensure responsible stewardship of trustworthy AI and more specifically inclusive growth, sustainable development and well-being, respect for the rule of law, human rights and democratic values including fairness and privacy, transparency and explainability, robustness, security and safety, accountability⁴⁵⁷.

2.5.3.6 Ethics by Design and Ethics of Use Approaches for Artificial Intelligence⁴⁵⁸

This document builds on HLEG Guidelines and offers guidance for adopting an ethically focused approach at the stage of design, development, deployment and use of AI based solutions using the approach of ethics-by-design.

Considered as suggested approach for Horizon Europe Applicants and beneficiaries, the guidelines define the 6 principles that AI systems should adhere (respect for human agency, privacy and data governance, fairness, individual, social and environmental well-being, transparency, accountability and oversight) and refer to practical steps on their application during the development of AI system defining 6 phases.

⁴⁵⁷ <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449>

⁴⁵⁸ European Commission DG Research & Innovation RTD.03.001- Research Ethics and Integrity Sector, Ethics by Design and Ethics of Use Approaches for Artificial Intelligence, v.1 25 November 2021, https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-by-design-and-ethics-of-use-approaches-for-artificial-intelligence_en.pdf

These are specifications of objectives, requirements, high-level design, data collection and preparation, detailed design and development, testing and evaluation of the system.

Finally, the ethical deployment and use is explained, and a check list regarding the specification of objectives against ethical requirements is included.

2.5.3.7 Guidance regarding Ethical AI use in the energy sector, OFGEM

The Guidance regarding Ethical AI use in the energy sector⁴⁵⁹, issued by OFGEM, the independent Regulator of Great Britain's energy markets, provides guidance with regards to safe AI, secure AI, fair AI and environmentally sustainable AI.

It further analyses good practices in governance and policies, risks and competencies. It also provides sector specific examples such as AI in consumer interactions, AI used to identify and assist excluded consumers, AI in predictions and forecasting, AI in cyber- physical systems, AI in pricing and trading and use of black boxes.

2.5.4 Relevance to EnergyGuard

Given the above, despite the significant benefits AI can offer across various sectors, attention must be paid to addressing the ethical challenges associated with it, ensuring that its integration is sustainable and socially responsible.

In addition to the above, it should also be mentioned that with regards to ethics, the Consortium will take under consideration and operate in accordance with the guidance provided by European Commission in Ethics⁴⁶⁰ and data protection, so that related issues will be addressed appropriately. With regards to data processing taking place by the Consortium within the project, data protection by design and by default will apply ensuring that to the extent possible appropriate technical and organizational measures to implement data protection principles and integrate the necessary safeguards for the protection of the rights and freedoms of data subjects⁴⁶¹. Consortium will also consider measures related to living guidelines on the responsible use of generative AI in research⁴⁶² (to the extent applicable) and thus is willing to be transparent with regards to generative AI tools to be used substantially in the research process and pay particular attention to issues related to privacy, confidentiality and intellectual property rights if and when sharing sensitive or protected information with AI tools.

⁴⁵⁹ <https://www.ofgem.gov.uk/sites/default/files/2025-05/ethical-ai-use-in-the-energy-sector-20250516150045.pdf>

⁴⁶⁰ European Commission, 5-7-2021, https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-and-data-protection_he_en.pdf

⁴⁶¹ European Data Protection Board, Guidelines 4/2019 on Article 25, Data Protection By Design and by Default, v 2.0, adopted 20-10-2020. Par. 7

⁴⁶² ERA Forum Stakeholders' document, Second version, April 2025, available at https://research-and-innovation.ec.europa.eu/document/download/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en?filename=ec_rtd_ai-guidelines.pdf

Regarding the EnergyGuard project, the following table presents an initial approach to the expected relevance of the aforementioned documents to the EnergyGuard project and the activities of its lifecycle.

No	Document	Category in relation to the project	Relevance to the project
1	Charter of Fundamental Rights for the European Union	General/ fundamental rights	It consists of the overarching fundamental rights framework that applies to the use of AI and AI Act must be interpreted in the light of the Charter
2	The Council of Europe Framework Convention on Artificial Intelligence	General/principles/regulatory sandboxes	General regulatory framework for AI systems
3	AI Act	General legislation applicable to AI systems	Horizontal regulatory framework regarding AI systems which is taken into consideration both for the operation of the EnergyGuard Platform to the extent applicable, as well as to examine whether AI applications to be tested in the TEF fall within its scope, to classify risk level, to define respective obligations and assess compliance of the AI application to be tested with regulatory requirements. AI Act rules also the operation of regulatory sandboxes, the testing of high risks systems in real world conditions outside AI regulatory sandboxes, as well as Union AI testing support structure
4	GDPR	Personal Data	Legal framework with regards to applicable provisions of GDPR.
5	Regulation 2018/1807 on a framework for the free flow of non-personal data in the European Union	Non personal data	When applicable, it can be used to assess legitimacy of free flow and sharing of non-personal data within the Union
6	Regulation (EU) 2023/2854 on harmonized rules on fair access to and use of data (Data Act)	Transfer of data	Related to transferring data. It focusses on facilitating data access and sharing between consumers, public bodies and businesses. Can be used to ensure that AI systems

			have diverse and high-quality data inputs
7	Directive 2002/58 (E-privacy Directive) as amended by Directive 2009/136	Personal data	Applicable as to cookies of the EnergyGuard Platform and to be used (if and to the extent applicable) as to the assessment of the legitimacy of the processing of data in AI applications to be tested in the EnergyGuard Platform
8	NIS 2 Directive	Cybersecurity	Subject to its transposition to national legislation of Member states, and to the extent applicable, it defines cybersecurity requirements for the covered entities (cybersecurity at the organizational level)
9	Regulation (EU) 2019/881 (Cybersecurity Act)	Cybersecurity	Regulatory framework that rules cybersecurity schemes related to cybersecurity requirement compliance of High-Risk AI systems
10	Regulation (EU) 2024/2847 (Cyber Resilience Act) (the main obligations of the Act will apply from 11 December 2027)	Cybersecurity	Defines cybersecurity requirements for high-risk AI systems subject to Cyber Resilience Act thus to be used as assessment to compliance with cybersecurity requirements when applicable (cybersecurity at a product level)
11	Directive 2019/790 (Digital Single Market Directive)	Copyright and related rights	Subject to transposition to national laws, when applicable (providers of general-purpose AI models) it refers to the legitimacy of data sources and can be used to examine compliance of the AI applications with article 53 (1) c of the AI Act regarding the copyright exception for text and data mining
12	Information Society Directive (2001/29)	Copyright	Subject to national transposition it applies as to copyright issues related legitimacy of data sources of AI systems/models
13	Directive 96/9 on the legal protection of databases	Copyright	Subject to national transposition it applies as to the legitimacy of the data sources of AI systems/models with regards to databases

14	Regulation 2022/2065 (Digital Services Act)	AI systems as intermediary services	Defines additional liability and obligations of intermediary services in case an AI system operates as such. N/A
15	European Declaration on Digital Rights and Principles	General/human Rights/ethics	General document regarding fundamental rights for digital transformation and use of new technologies
16	Approval of the content of the draft Communication from the Commission - Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act), European Commission, 6-2-2025	Guidance on AI systems	Guidance for the assessment as to whether AI applications to be tested in EnergyGuard TEF constitute an AI system within the meaning of the AI Act as to the criterion of AI system, to enable further assessment of the application of the AI Act
17	Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models	Personal Data	Guidance as to whether an AI model can be considered as anonymous to be used for the assessment of the applications to be tested
18	HUDERIA methodology	Fundamental rights impact assessment	Guidance for the conduct of fundamental rights impact assessment, when necessary, by AI Act
19	Guidelines on prohibited artificial practices established by Regulation (EU) 2024/1689 (AI Act), European Commission, 4-2-2025	AI Act Prohibited Practices Guidelines	Non-binding guidance to ensure compliance with obligations with AI Act with regards to prohibited practices
20	OECD Recommendation on Artificial Intelligence	General/intergovernmental standard/regulatory sandboxes/ethics	It includes general recommendations to governments for the development of regulatory sandboxes. In addition, as first intergovernmental standard on AI, it is also important due to its supranational character containing

			five high level values-based principles and five recommendations for national policies and international cooperation
21	The Commission Recommendation C (2022) 3219 on speeding up permit-granting procedures for renewable energy projects and facilitating Power Purchase Agreements	Recommendation for the development of regulatory sandboxes in energy sector	Underlines the importance of establishment of regulatory sandboxes the energy sector to support innovation
22	Directive (EU) 2023/2413	Energy/regulatory sandboxes	Subject to transportation to national technologies, Member states should promote innovative energy technology through pilot projects in a real-world environment
23	Ethical Guidelines for trustworthy AI, High-Level Expert Group on Artificial Intelligence, and (HLEG Guidelines) Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment	Ethics	The Guidelines and List can be used as a tool for the assessment of AI applications to be tested in EnergyGuard TEF with regards to the achievement of trustworthy AI and offer guidance on how to secure the development of ethical AI systems in the EU. Also of importance for AI solutions developed within the project
24	Recommendation on the Ethics of Artificial Intelligence, UNESCO	Ethics	Ethical guidance for the assessment of AI applications to be tested in EnergyGuard TEF and for the inclusion of ethics in all stages of the AI system life cycle
25	European Commission, Ethics and Data protection	Ethics	Guidance with regards to handling of data protection and ethics issues raised during the project
26	Guidance regarding Ethical AI use in the energy sector, OFGEM	Ethics	Guidance and provision of good practices with regarding to the use of AI in the energy sector
27	Guidelines 4/2019 on Article 25 Data	Data protection	Guidance to be used for the implementation of data protection by

	Protection by Design and By Default		design and by default wherever applicable
28	Living Guidelines on the responsible use of generative AI in research	Responsible use of Generative AI in research	Non-binding directions on the responsible use of generative AI in research

Table 2: Expected relevance of the documents to EnergyGuard project

3 Requirements analysis

This section provides a comprehensive analysis of the user requirements that will guide the development of the EnergyGuard Testing and Experimentation Facility (TEF). Building upon the technical and regulatory insights from the previous sections, the requirements analysis aims to capture the needs, expectations, and priorities of a wide range of stakeholders across the AI and energy ecosystems. The data collection process included the design and distribution of targeted questionnaires (presented in Annex 2) and in-depth engagement with pilot use case leaders.

Section 3 is structured into three main parts:

- Section 3.1 details the inputs gathered from eight stakeholder categories, including AI developers, energy system actors, researchers, legal experts, and the general public.
- Section 3.2 focuses on the five pilot use cases, extracting functional and technical specifications for real-world validation
- Section 3.3 synthesizes these insights into a set of TEF specifications, including user personas, user stories, and high-level requirements.

3.1 Questionnaires input

This section presents the results of a series of targeted questionnaires conducted among key stakeholder groups to gather user needs and expectations relevant to the design of the EnergyGuard Testing and Experimentation Facility (TEF).

The questionnaires, which are included in full in Annex 2, were distributed across eight stakeholder categories—ranging from AI developers to legal experts and the general public—to ensure a comprehensive and inclusive requirements analysis. Section 3.1 is structured according to these stakeholder types, with each subsection summarizing the specific feedback, insights, and priorities identified for that group.

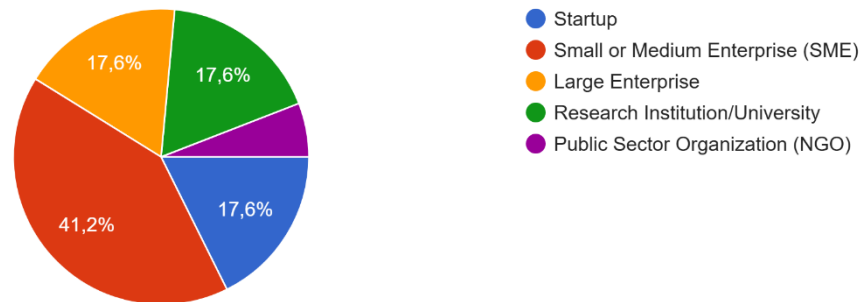
3.1.1 AI Developers

The questionnaire directed at AI developers received responses from stakeholders from different countries: Italy, France, Greece, Latvia, Lithuania, Portugal, Slovakia, Spain, and Ukraine

Most responses come from SMEs (41%), but startups, large enterprises, research institutions, and NGOs are also represented:

Type of organization:

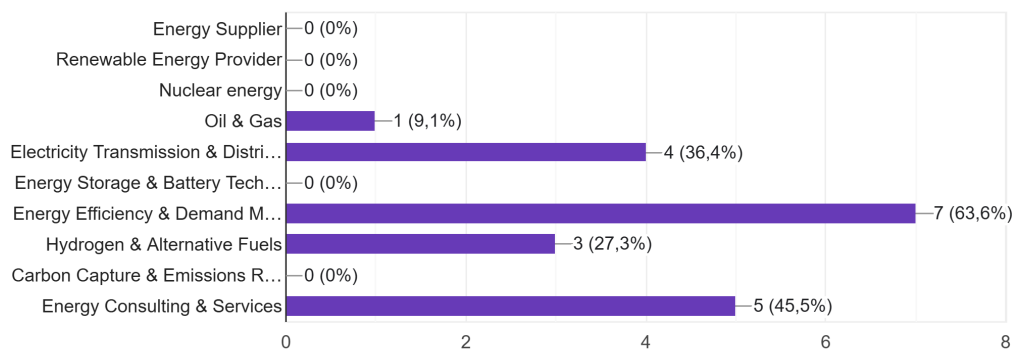
17 risposte



The industry sector that received the most selections is 'Energy efficiency and demand management'. Other options that received several preferences are 'Energy consulting and services' and 'Electricity transmission and distribution'.

Energy

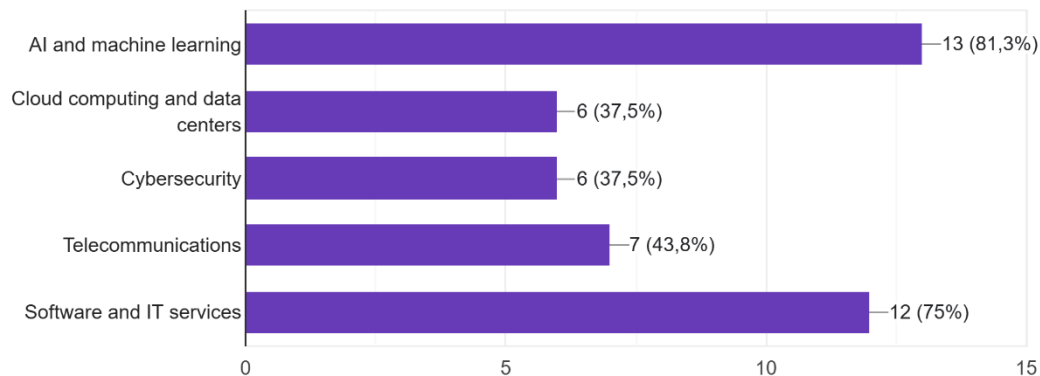
11 risposte



The respondents work in various ICT domains. Most of them selected the options 'AI and machine learning' and 'Software and IT services':

ICT

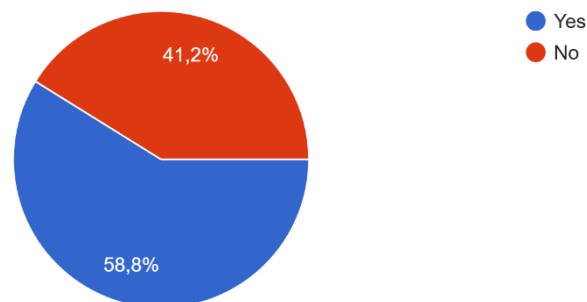
16 response



The majority of respondents developed or plan to develop AI applications for the energy sector:

Do you develop or plan to develop an AI application for the energy sector?:

17 response



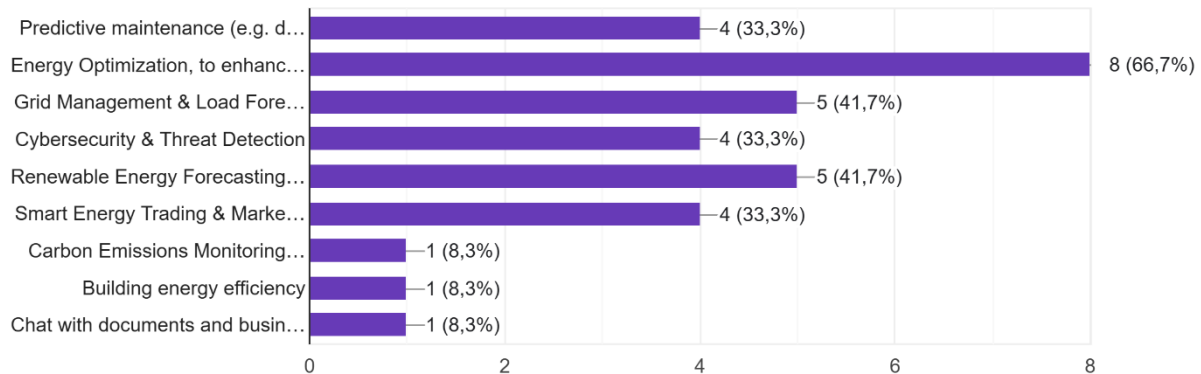
Respondents reported various AI applications in the energy sector, including automated energy performance certificates (EPCs) and retrofit recommendations, predictive maintenance of industrial machinery using AI-based monitoring, and forecasting and optimization solutions through machine learning. Some projects focus on large-scale building energy assessment and retrofit planning, as well as the creation and management of energy communities supported by AI.

Other initiatives involve digital twins of transmission grids, AI-based fault detection and classification in power systems, and voltage profile estimation using AI rather than grid models. Additional applications include grid frequency monitoring, right-of-way (RoW) management, and tools aimed at reducing energy costs in energy-intensive industries.

The main objective of these AI solutions seems to be energy optimization, to enhance efficiency in power generation, distribution and consumption:

What is the main objective of your AI solution?

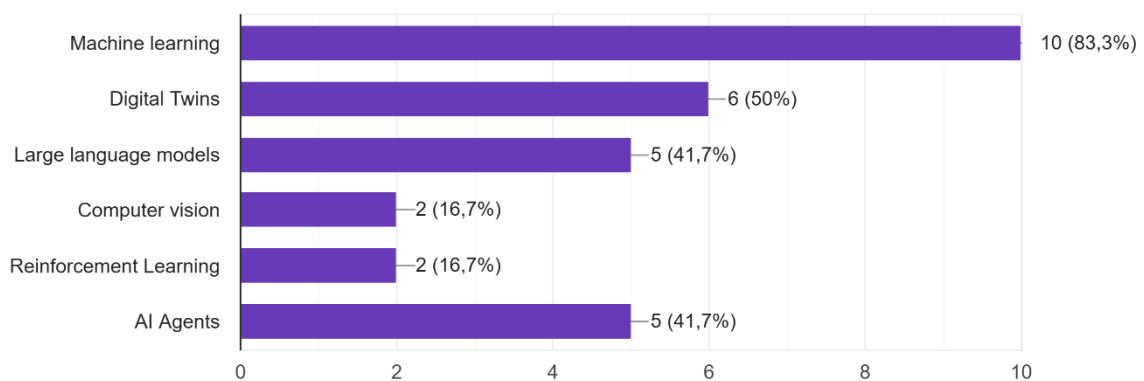
12 risposte



Among AI technologies, the most widely employed are machine learning, digital twins, LLMs and AI agents:

What AI technologies/methods are you using?

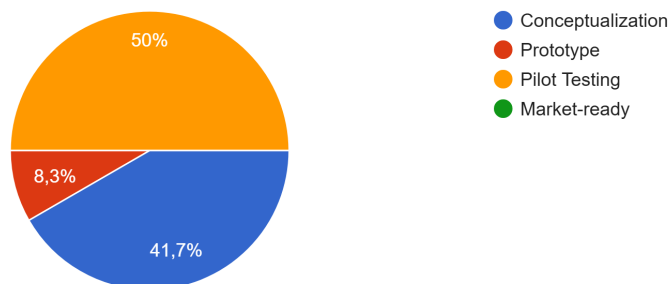
12 risposte



Most AI applications are in pilot testing phase, and many are only in conceptualization phase:

What stage of development is your AI solution currently in?

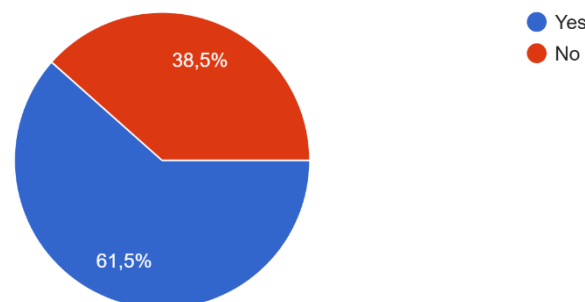
12 risposte



Respondents highlighted AI applications tested or deployed in diverse real-world contexts. These include projects carried out through the I-ENERGY initiative and implementations in various industrial sectors such as oil and gas, food processing, mining, metallurgy, and manufacturing. AI solutions have been applied in solar plants, residential retrofitting (with pilot projects in collaboration with VINCI in France), and energy communities across Italy. Use cases also involve real-time testing of automatic fault analysis in physical power system labs, and deployment in facilities like conference centers, vegetable warehouses, and bakeries. Reported outcomes include improved accuracy, reduced evaluation time, and significant cost savings.

Have you previously tested your AI solution in a real-world environment?

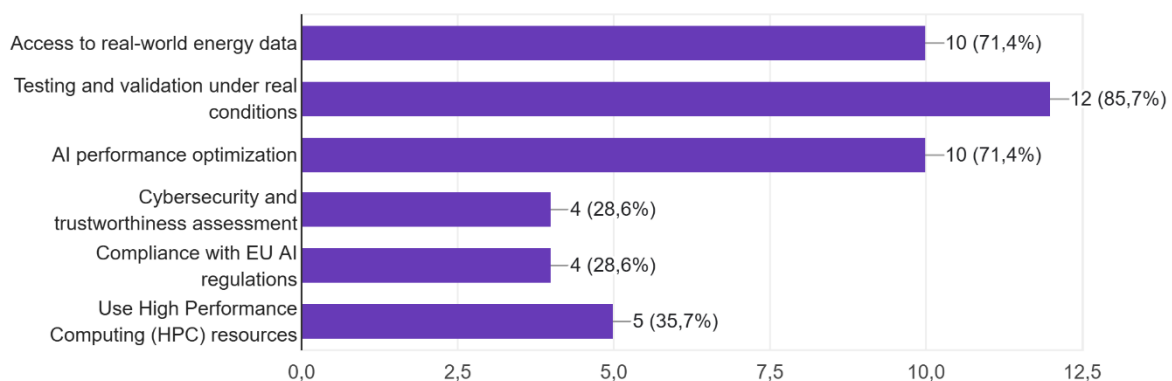
13 risposte



When asked about their main objectives for using the EnergyGuard TEF, respondents selected several options. The most common were testing and validation under real conditions, access to real-world data, and AI performance optimization:

What are your main objectives for using the EnergyGuard TEF?

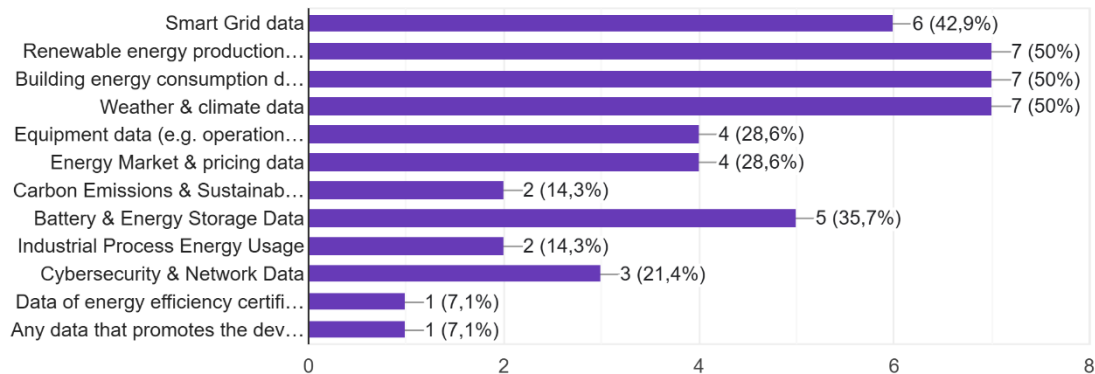
14 risposte



The answers to the question about what type of data is needed for testing reveal a variety of needs, suggesting that all types of data are relevant:

What type of data would you need for your testing?

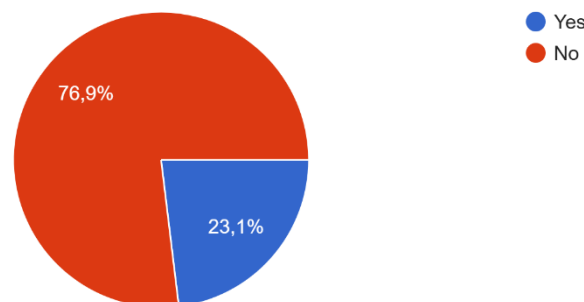
14 response



Quite unexpectedly, regulatory and compliance requirements do not appear to be a major concern at this stage. Respondents who expressed the need to meet regulatory requirements mentioned European Union's AI Act, GDPR and NIS 2 Directive (Network and Information Security).

Are there specific AI regulatory or compliance requirements that you need to meet?

13 response



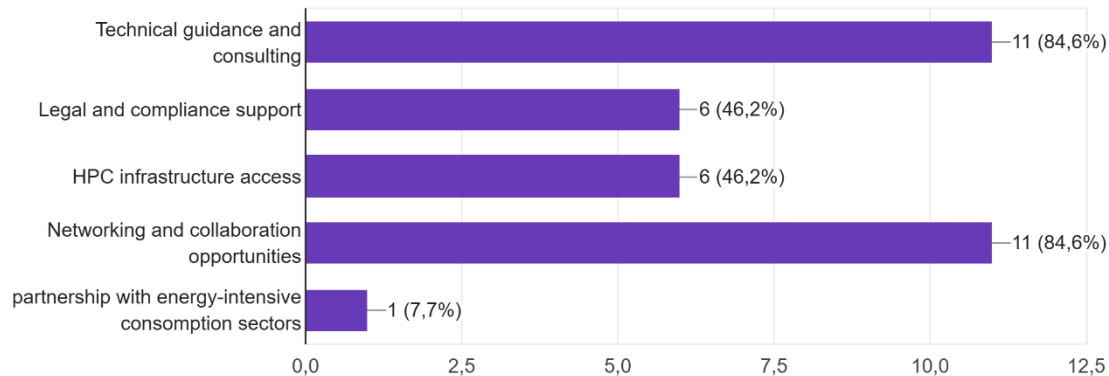
When asked about expectations regarding EnergyGuard TEF's web user interface, users responded that they expect the interface to allow easy viewing of available data and metadata, access to datasets (including downloading and API access), and exploration of individual datasets. Some want to test and access their own applications. Priorities include data availability, resource search functionality, and viewing facilities.

When asked about expectations regarding EnergyGuard TEF's APIs, users responded that they expect EnergyGuard TEF's APIs to offer high performance and support large datasets (in the terabyte range), with fast data ingestion and automated trustworthiness checks. Security, ease of use, and compatibility with tools like Python are also important, as is the availability of manuals or tutorials. Some users emphasize flexibility in choosing dataset sizes and the importance of data reliability.

The expectations regarding support services emphasized technical guidance and consulting, as well as networking and collaboration opportunities:

What are your expectations regarding EnergyGuard TEF's support services?

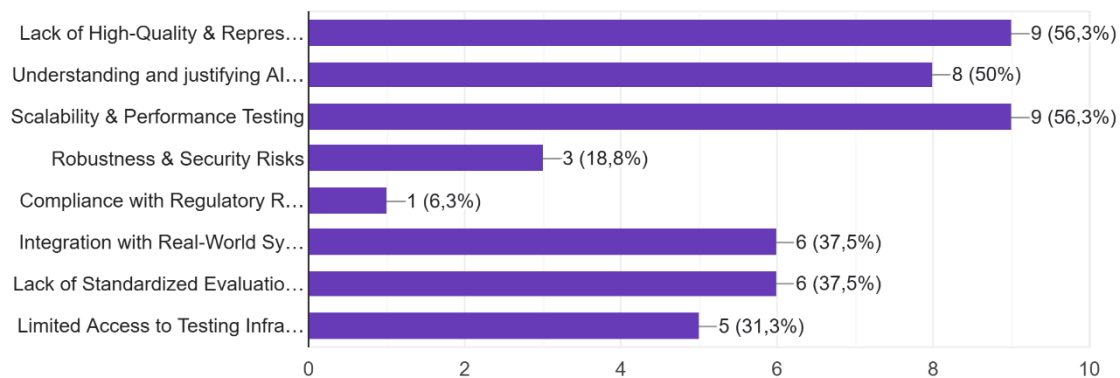
13 risposte



The biggest challenges faced in AI applications testing and validation are lack of high-quality data, scalability, performance, and understanding AI decisions:

What are the biggest challenges you face in AI testing and validation?

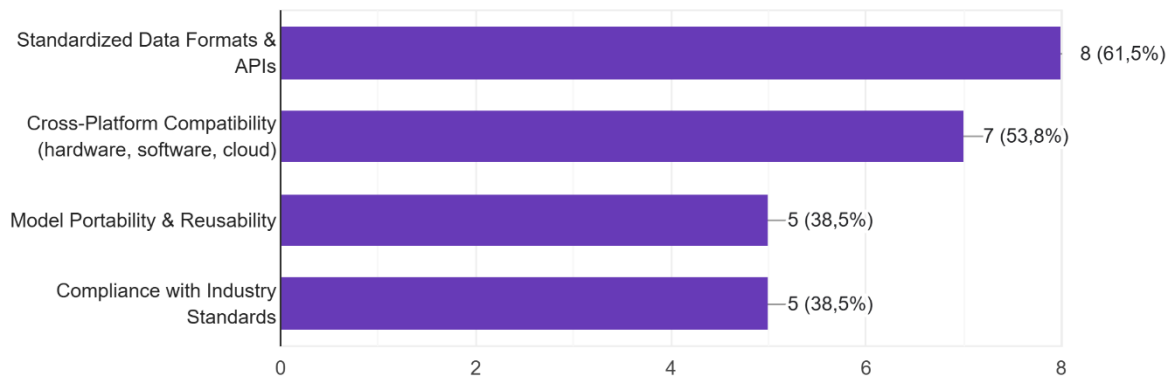
16 risposte



The requirements for interoperability are in favour of standardized data formats and APIs and cross-platform compatibility:

Do you have any specific requirements for AI model interoperability?

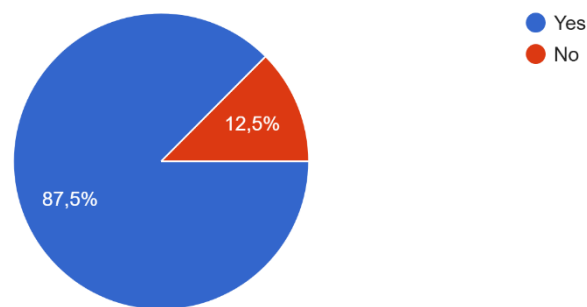
13 risposte



Finally, most respondents say that they are interested in collaborations:

Are you interested in collaborating with other AI developers, researchers, or energy sector stakeholders?

16 risposte

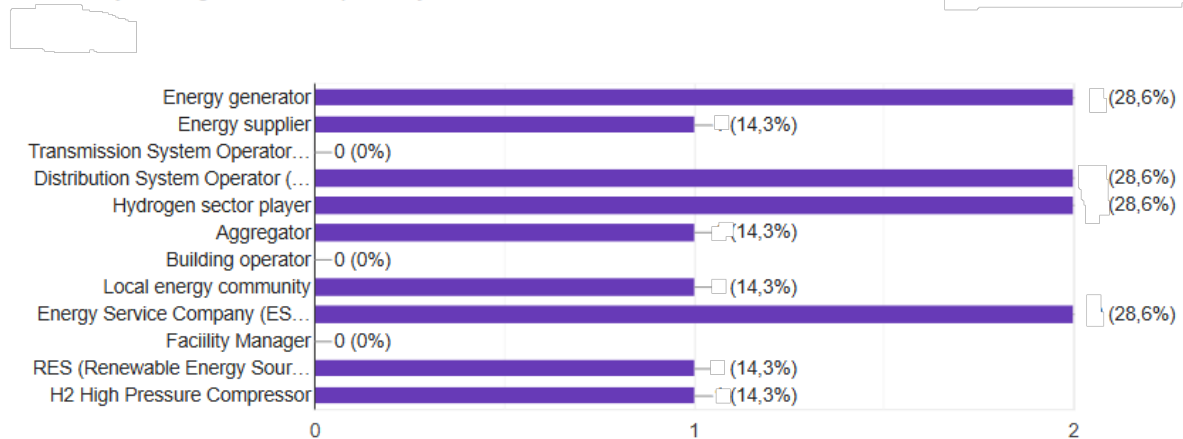


3.1.2 Energy Production and Energy Systems (EPES) value chain stakeholders

The questionnaire directed at organizations involved in energy production and energy systems received responses from different countries, namely Spain, Belgium, Italy, and the USA.

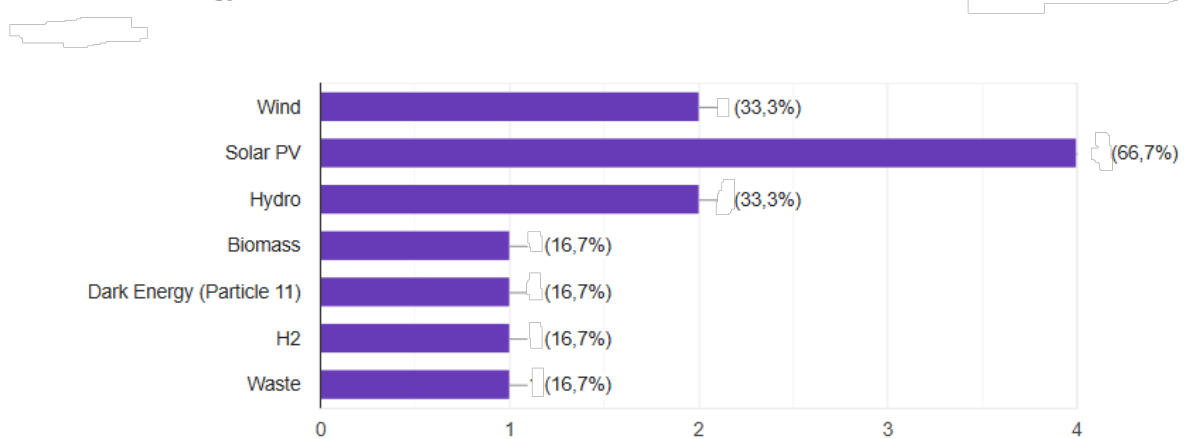
These organizations identified their primary roles in different ways:

What is your organization's primary role in the EPES sector?



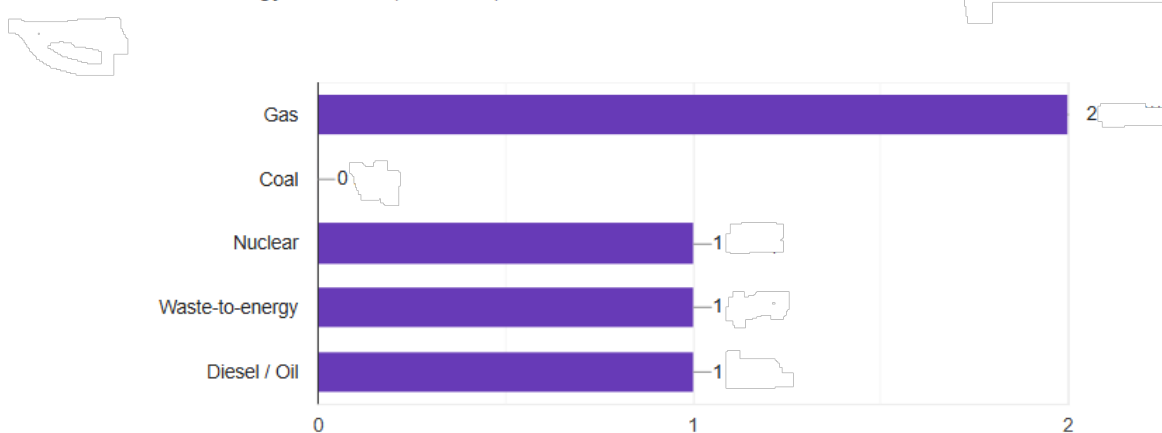
Most organization work with solar PV as primary renewable energy source:

Renewable energy sources:

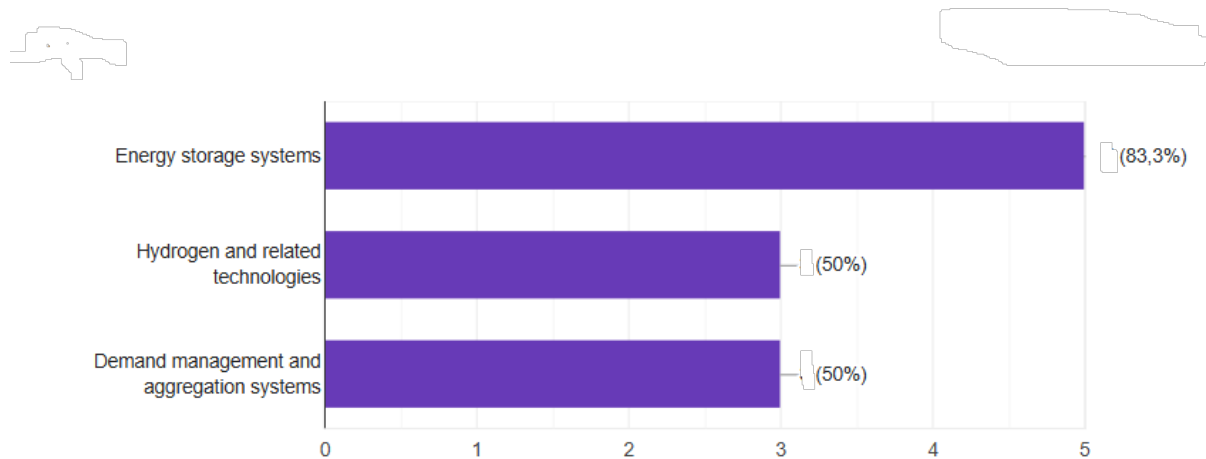


Some organization also work with non-renewable energy sources:

Non-renewable energy sources (non-RES):

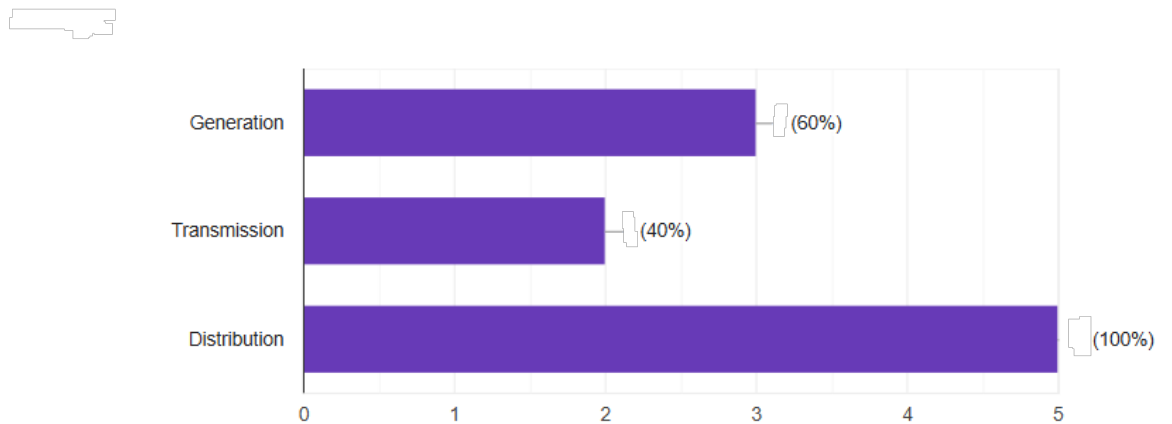


The energy technology that these organizations are primarily responsible for are:



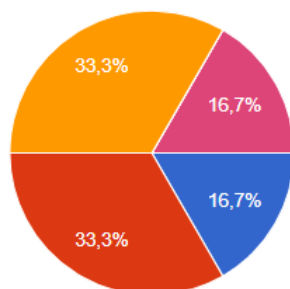
All respondents selected 'energy distribution' as their primary responsibility, but also mentioned energy generation and transmission:

Energy grid technologies:



Most respondents identified AI for predictive maintenance and AI for demand-side management and the most promising technologies that they are willing to adopt:

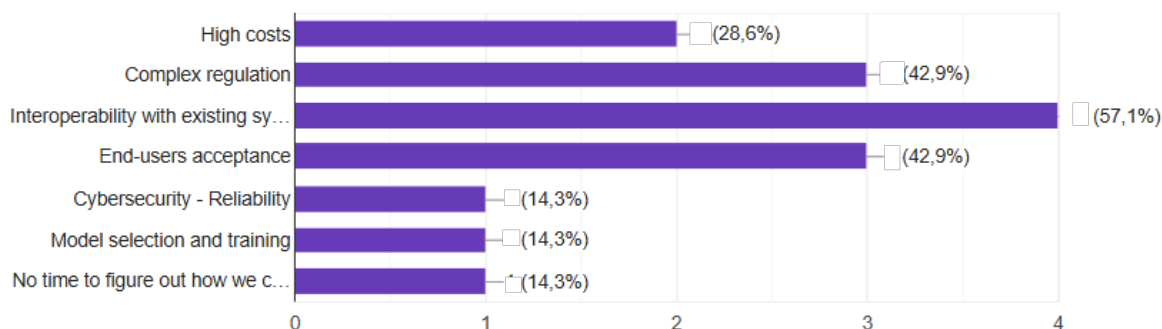
What are the new solutions and innovative technologies that you are willing to adopt or are interested in, and what are the main challenges your organization faces in adopting them?



- AI for forecasting and optimization
- AI for predictive maintenance and asset management
- AI for demand-side management and customer interaction
- AI for operational security and anomaly detection
- AI-based digital twins and simulation t...
- AI for Distributed Energy Resources a...
- All of the above

The main identified barrier to adopting these innovative technologies is interoperability with existing systems:

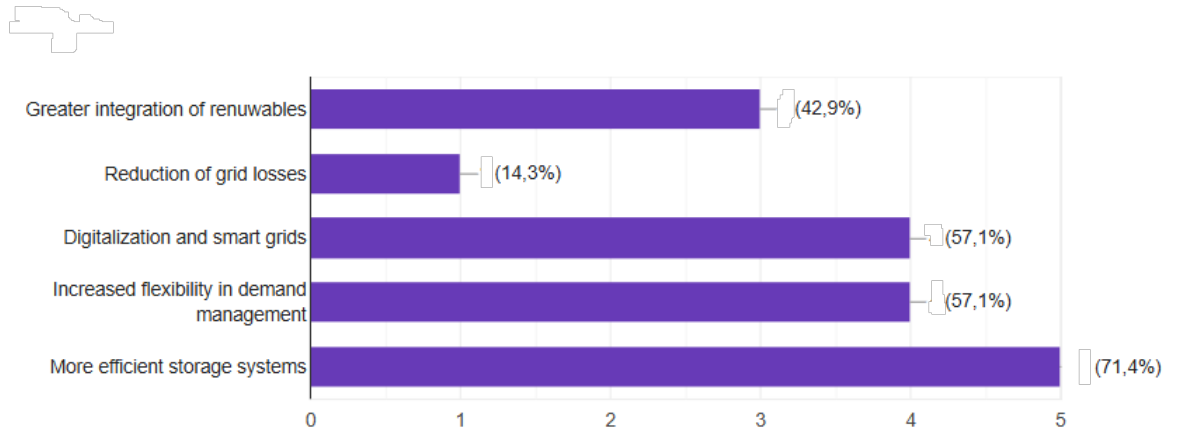
What are the key barriers to adopting these innovative Technologies in the EPES sector? (Select up to 3 options)



When asked about the needs to improve operational efficiency and sustainability, organizations highlighted the need for real-time data access, predictive maintenance, and integration of diverse data sources to improve operational efficiency and sustainability. Key priorities also include advanced forecasting tools, AI-driven decision support systems, enhanced monitoring and control of low-voltage networks, collaboration at the EU level, and flexible regulatory frameworks that support innovation and scalability.

The most important priority for developing new solutions and innovative technologies is the availability of efficient systems for energy storage:

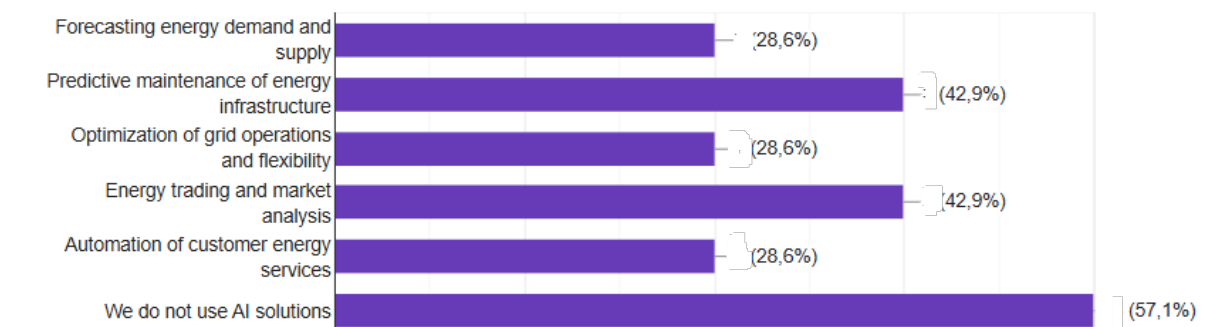
What aspects do you consider a priority in developing new solutions and innovative technologies? (Select up to 3 options)



Organizations use a variety of digital tools and platforms to optimize energy, asset, and system management. These include IoT platforms (e.g., ThingsBoard), smart grid management systems, and AI-based solutions for predictive maintenance and anomaly detection. Digital Twins are being developed to simulate grid and asset behaviour. Other tools mentioned are forecasting and flexibility optimization platforms, custom dashboards for visualization and control, and data integration platforms ensuring secure interoperability. Some organizations also rely on in-house Python libraries or partner solutions, while one respondent reported not using any digital tools now.

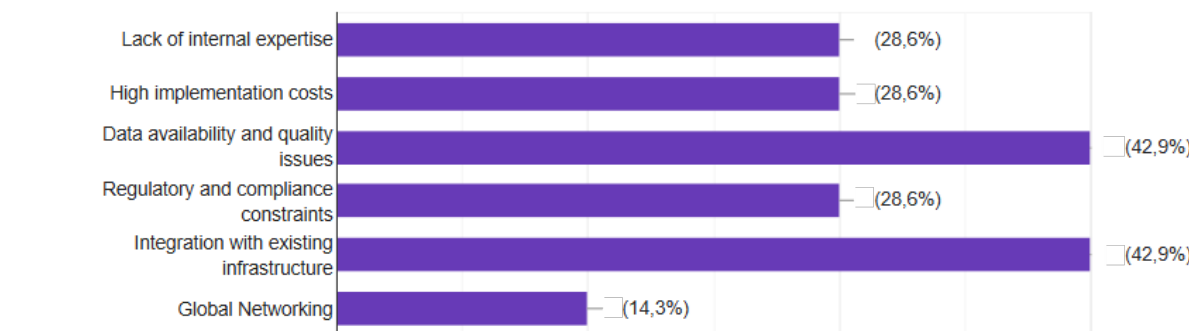
Some organizations already employ AI solutions, but most don't:

Does your organization currently use artificial intelligence (AI) solutions in energy/asset/system management, and in which areas?



The main faced challenges in adopting AI-based solutions are availability of quality data and, again, possible integration problems with existing infrastructures:

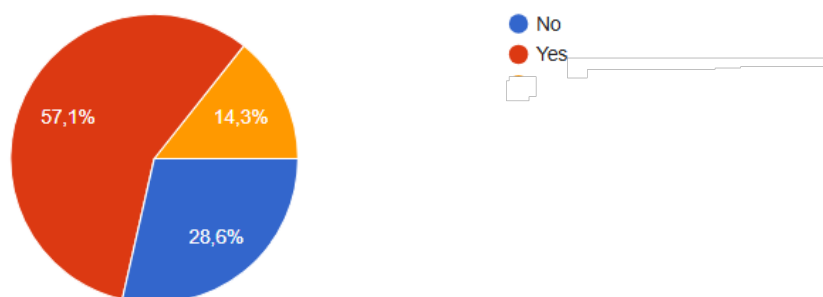
What are the main challenges your organization faces in adopting AI-based solutions for energy systems? (Select up to 3 options)



The most impactful regulations identified by organizations include EU-level policies such as the EU AI Act, the Clean Energy Package, and various EU energy directives. Key areas of regulatory importance are consumer access to electricity markets, data privacy and cybersecurity (e.g., GDPR, NIS 2), and national/regional rules for DSOs, particularly in Spain. Respondents also emphasized the significance of flexibility market regulations, including those that support small-scale asset participation, as well as incentives and obligations related to renewable energy, self-consumption, and energy communities.

Most organizations have specific decarbonization or sustainability goals:

Does your organization have specific decarbonization or sustainability goals?



To facilitate the energy transition, organizations call for clearer regulatory frameworks for flexibility and local energy markets, enabling DSOs to act as neutral facilitators. Desired support includes subsidies for energy off-takers, incentives for energy communities, simplified administrative procedures, and funding for grid innovation and digitalization. There is also interest in regulatory sandboxes to test new solutions, as well as incentives for technology and AI adoption and support for cybersecurity compliance. Additionally, enabling direct consumer access to power exchanges and better matchmaking mechanisms were mentioned.

To improve collaboration across the EPES value chain, respondents suggest standardizing data protocols and interfaces to ensure interoperability and promoting shared digital platforms and data spaces for secure, transparent information exchange. Other key recommendations include inclusive regulatory frameworks that support smaller actors, joint pilot and demonstration projects to align goals and build trust, and capacity building and common training to foster shared understanding and

innovation adoption. One respondent emphasized optimizing the entire value chain from production to consumption, while another reported having no specific suggestions.

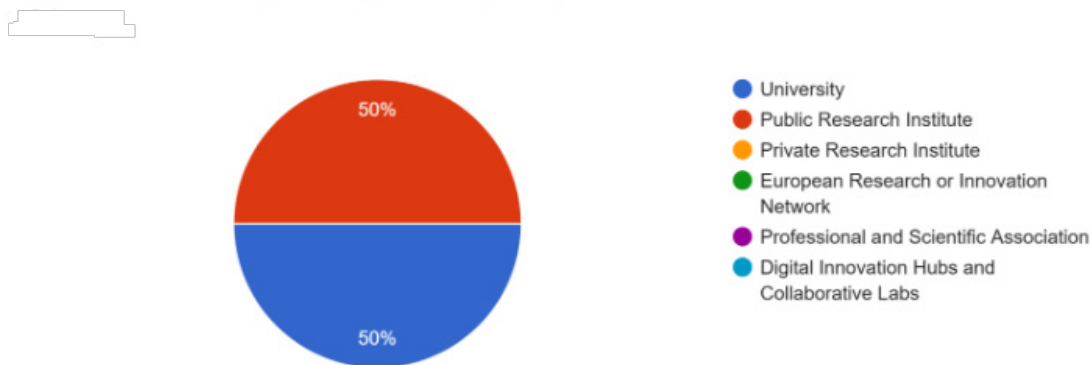
Additional comments highlight the need for regulatory flexibility to allow DSOs a more proactive role in managing distributed flexibility. Respondents emphasized cybersecurity and data governance as key challenges in deploying AI and digital tools. There's also a call for support in scaling successful pilot projects, which often stall due to regulatory and market barriers. One organization expressed interest in aligning a consumer-centric use case with the project, and another welcomed collaborative European initiative that integrates technical, regulatory, and social aspects of the energy transition.

3.1.3 Researchers and Academia

The questionnaire directed at researchers and academia received responses from stakeholders from different countries: Greece, Italy, Latvia, Spain, and United Kingdom.

Half of the respondents identified themselves as Public Research Institutes, and half as universities.

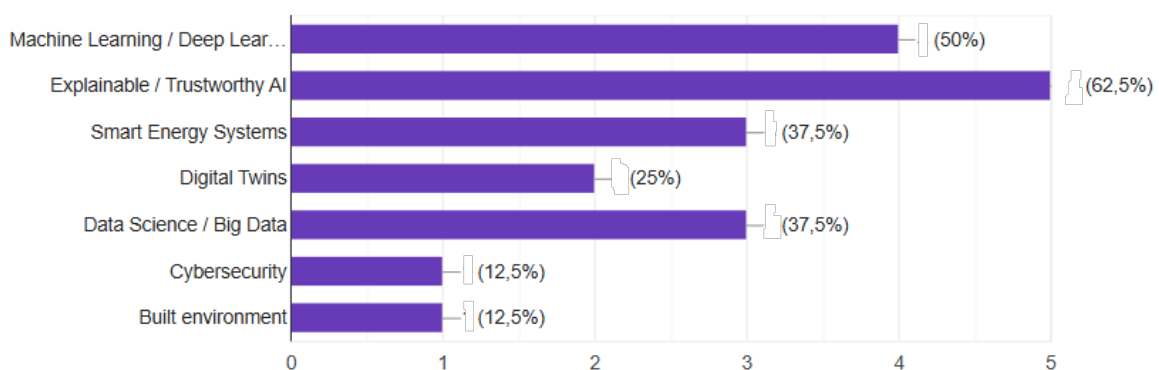
Please indicate the type of organization you represent:



The main areas of research are explainable/trustworthy AI, machine/deep learning, smart energy systems and big data:

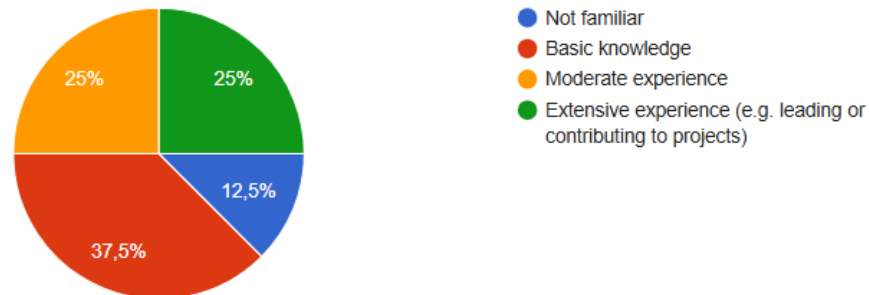
What is your main area of research? (Select one or more)

8 risposte



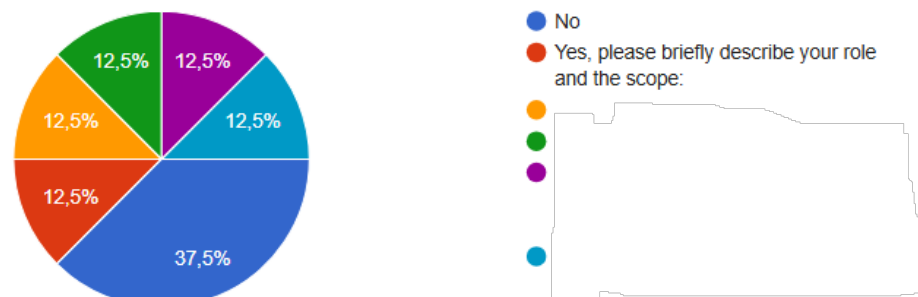
The level of experience and familiarity with applications of AI and Machine Learning in the energy domain does not indicate a predominant type of answer:

How familiar are you with applying AI/ML in the energy domain?



Most respondents say they are currently involved in research projects related to AI in energy systems (the different colours represent the different descriptions of project types).

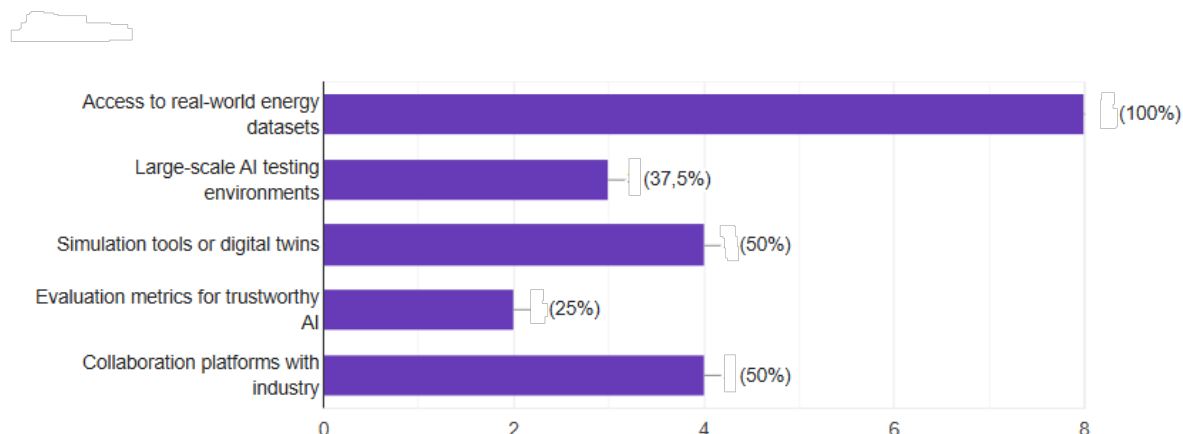
Are you currently involved in any research projects related to AI in energy systems?



Several topics were identified as priorities for further research to support the trustworthy adoption of AI in energy systems. These include the development of compliance frameworks and standardised trustworthiness assessment methods, ensuring data quality and governance, and improving the explainability and interpretability of AI models. Other key areas are integration with human decision-making, scalability, energy efficiency of AI solutions, interoperability, and the standardisation of technologies like digital twins. Robustness under uncertain conditions and reproducibility of AI outcomes were also highlighted as essential for safe and effective deployment.

The type of support or infrastructure which is considered as the most useful is access to real-world energy datasets:

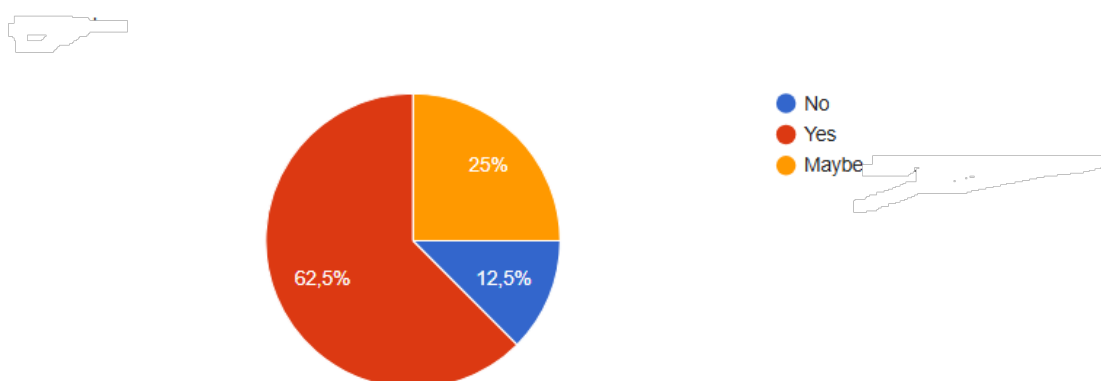
What types of research support or infrastructure would be most useful for your work? (Select up to 3)



Several challenges hinder the transfer of research into real-world energy applications. Key issues include limited availability and poor quality of real-world data, lack of access to domain expertise, and difficulty in scaling solutions from laboratory to industrial environments. There is also reluctance from operators and stakeholders to adopt unproven AI systems due to trust and safety concerns. Additional barriers involve integration with legacy infrastructure, lack of standardisation and interoperability, regulatory compliance, and insufficient collaboration across disciplines. In some cases, technology transfer is simply not a primary research focus.

Most respondents say they are interested in contributing to EnergyGuard through scientific collaboration:

Would you be interested in contributing to EnergyGuard through scientific collaboration (e.g. piloting, testing, co-authoring papers, etc.)?



To ensure alignment between the EnergyGuard infrastructure and academic research needs, respondents recommend facilitating access to high-quality, well-documented datasets and providing real-world testing environments. Transparent communication and early involvement of researchers—through open dialogue, workshops, and collaborative projects—are considered essential. Other suggestions include promoting open-access results, ensuring flexibility and interoperability of the platform, and clearly highlighting EnergyGuard's unique value compared to commercial cloud services, such as access to exclusive data or lower costs.

3.1.4 Other stakeholders

Other questionnaires were directed to industry associations, technology clusters, investors, policy makers, standardization organizations, lawyers, legal system entities, and the general public. These questionnaires received answers from stakeholders from Italy, Belgium, Latvia, and Greece.

Here is a summary of most important feedback from these questionnaires:

- Most critical risks in deploying AI in the energy sector are data protection and privacy issues, as well as lack of transparency in AI decision-making.
- It is very important to know that AI systems in energy are tested for safety, transparency, and fairness before being used.
- Main barriers to adopting AI in energy systems are data availability and interoperability and fragmentation of standards and platforms.
- The areas that should be prioritized in future AI testing and experimentation for the energy sector are energy communities and prosumer engagement.
- The most impacted areas of policy, regulation or standardisation in energy systems are data protection, cybersecurity, consumer rights, certification and standardisation of AI tools.
- The main regulatory or standardisation gaps involve cross-border standardization, regulatory clarity on liability and risk allocation, lack of security and cybersecurity standards, certification and transparency of AI systems.
- There should be convergence among regulation and standardization. It is necessary that all compliance requirements provided for in the AI Act are also implemented in the development of standardization or specifications in the energy sector. It is also necessary to develop possible working groups to communicate with the AI Board or Office on these issues and develop a unified, common approach.
- To ensure trustworthy AI in the energy sector, guidelines from competent authorities are needed, as well as sector-specific regulatory sandboxes.
- There's interest in receiving updates about new tools or services developed through EnergyGuard, preferably via website and newsletters.

3.2 Pilot use cases

The following sub-sections report the results from the questionnaire directed to partners responsible for pilot applications, that are the experiments to run or the apps to develop with the use of the TEF facilities. For each pilot application, one or more use cases are listed.

3.2.1 Pilot 1 – RDN

This pilot consists of 3 different use cases, listed in the following sub-sections.

3.2.1.1 Use Case 1 – Digital Twin of the Portuguese Transmission System Operation System

Overview	
Owner of the Pilot	R&D NESTER
Use Case Acronym	P1_UC1
Document Version	V1.00
Scope / Objectives of Use Case-Experiment	

This use case focuses on the integration of the DTs of the large-scale Portuguese Transmission Network and Power simulations with the EnergyGuard development testing sandbox environment. It includes the ingestion and processing of real-time and historical data from the Transmission System Operator (TSO) into the EnergyGuard data lake, covering electric grid operations and electricity demand across multiple geographic regions.	
Primary Actor	The System Operator (TSO)
Secondary Actors	AI developers
Equipment / Hardware used	Use of the RTPSS Laboratory
Use Case Description	
In the envisioned future, System Operators (TSOs) and AI developers will collaboratively leverage the Digital Twin (DT) of the Portuguese Transmission Network within the EnergyGuard sandbox to test and validate AI-driven grid optimization models.	
User Perspective	
From the user's perspective particularly for AI developers and grid operators, the platform offers a major leap forward from the current state of AI testing in the energy domain. Currently, validating AI models for smart grid optimization is limited by data access restrictions, regulatory uncertainty, and the lack of realistic test environments. The service I TEF provides a secure access to a large-scale Digital Twin of the Portuguese Transmission Grid for realistic and high-fidelity experimentation.	
System Perspective	
From a system perspective, the TEF platform interacts with existing grid infrastructure by integrating with Digital Twins, SCADA systems, and data sources from TSOs. It enables secure data exchange, simulation-based testing, and real-time feedback between operational platforms and AI models.	
Details	
Business Goals/Benefits	
This use case is executed to address critical challenges in the energy sector by leveraging Digital Twins (DTs) and AI-driven grid optimization within a controlled sandbox environment. • Enable safe testing of AI/ML models (e.g., demand forecasting, fault detection) without risking real-grid stability. • Reduce barriers to innovation caused by data privacy regulations or lack of test environments.	
Application	• Enables real-time grid emulation for validating AI models under dynamic conditions. • A secure testing platform for AI/ML models in energy systems.
Data Preconditions	Historical grid operational data from the TSO (e.g., load profiles, voltage/current measurements, grid topology, fault records).
System Preconditions	Fully operational testing environment with access controls and user authentication.

User Preconditions	Users (TSO operators/AI developers) must complete platform registration and TBD.
Trigger	AI developer completes prototype phase and requests real-world validation.
Challenges/Limitations	Cybersecurity and data privacy concerns during data transmission and processing. Requirement for validation of AI decisions before full automation is trusted in critical operations.
Expected results - KPIs	
Execution time: Real-time transitory emulation of a 1000-bus transmission grid's behaviour between 15-min market periods is expected to take less than 3 minutes.	
User Journeys	
- A user wants to test how the grid behaves (real-time) between two different 15 –min market results. - The user submits 2 different grid states, i.e., the expected steady-state power flow (either 2 .csv files or 1 file with both states labelled). - The platform ingests and triggers an automated analysis. - Within minutes, a report is available to the user via .csv file with the calculated steady-state power flow results during both states and some transient period power flow alarms (exceeded bus max/min voltage, max/min power, etc.).	

3.2.1.2 Use Case 2 – Automated Fault Analysis of Power System Events

Overview	
Owner of the Pilot	R&D NESTER
Use Case Acronym	P1_UC2
Document Version	V1.00
Scope / Objectives of Use Case-Experiment	
Service II will focus on automatic fault analysis of power system events. The service leverages signal processing technique and AI tools for detecting and classifying the faults, as well as quantifying event's statistics (e.g. detection time, short-circuit current). It can automatically process the COMTRADE files generated by Intelligent Electronic Devices (IED) such as protective relays, providing incident reports within a few tens of seconds from the occurrence of an event.	
Primary Actor	The main actor is the System Operator who is responsible for monitoring and managing the electrical grid and requires rapid incident analysis to support operational decisions.
Secondary Actors	AI developers

AI techniques / Models	Usage of AI techniques: ML models such as Decision trees, eXtreme Gradient Boosting
Equipment / Hardware used	Use of the RTPSS Laboratory
Software used	Python
Use Case Description	
Ensuring accurate fault classification across diverse scenarios and noise conditions. Achieving low-latency processing for real-time response. Handling large volumes of COMTRADE files from distributed sources. Validating performance in realistic grid scenarios using the Digital Twin.	
User Perspective	
In the envisioned future, fault analysis will be fully automated and integrated within real-time grid operations. When a fault occurs, the IEDs will send COMTRADE files to a centralized platform. The system will instantly process these files, identify the fault type and location, calculate key statistics, and generate incident reports. These reports will be available to system operators, protection engineers, and maintenance teams within seconds, enabling faster decision-making and reducing outage times. Testing this workflow in the Digital Twin will ensure reliability, accuracy, and seamless integration before deploying it in the live system. From the user's perspective, the TEF enables secure, realistic testing of the automatic fault analysis service without risking real-world assets. It allows them to assess the system's performance under various fault scenarios, verify its accuracy, and evaluate how it integrates with existing operational tools.	
System Perspective	
From a system point of view, the fault analysis platform is designed to be modular and interoperable. It can be connected to SCADA systems, protection databases, and asset management platforms. Within the TEF, it interacts with the Digital Twin by receiving simulated fault event data as input and returning processed fault reports. This loop allows the platform's behaviour to be studied and refined in controlled yet dynamic conditions, ensuring it can later be deployed in the real grid with minimal integration friction.	
Details	
Business Goals/Benefits	
The goal of this use case is to enable rapid and automated fault analysis in the transmission network using AI and signal processing. The benefits include: • Faster identification and classification of faults • Reduced system downtime and quicker restoration of service • Improved accuracy in fault diagnostics • Enhanced situational awareness for operators • Reduced manual workload and human error • Better planning and decision-making through detailed incident reports	

Application	The use case pertains to a fault analysis application. It processes COMTRADE files and provides real-time fault analytics and reporting.
Data Preconditions	COMTRADE files must be generated and available from IEDs (e.g., protective relays) after a power system event.
System Preconditions	The fault analysis platform must be properly deployed and connected to receive data from the DT or live IEDs.
User Preconditions	Users should be trained to interpret the output reports or interact with the analysis dashboard .
Trigger	The use case is triggered when a fault event occurs in the transmission network and a corresponding COMTRADE file is generated by an IED and sent to the analysis system.
Challenges/Limitations	AI model accuracy depends on training data quality and representativeness. Cybersecurity and data privacy concerns during data transmission and processing. Requirement for validation of AI decisions before full automation is trusted in critical operations.
Expected results - KPIs	
<u>Accuracy</u>: Fault classification accuracy is targeted to exceed 90% for the most common fault types. <u>Execution time</u>: Incident reports are expected to be generated within 30 to 60 seconds after receiving the COMTRADE file.	
User Journeys	
User Journey: - A fault occurs on the grid. - IED captures and sends a COMTRADE file. - The platform ingests the file and triggers automated analysis. - Within seconds, an incident report is available to the operator via dashboard or notification. - Protection engineers receive detailed data for post-event validation or corrective action.	

3.2.1.3 Use Case 3 – Model Free Voltage Estimation

Overview	
Owner of the Pilot	R&D NESTER
Use Case Acronym	P1_UC2
Document Version	V0.0

Scope / Objectives of Use Case-Experiment	
Service III will focus on model-free voltage profile estimation of Low-Voltage (LV) networks. Detailed electrical grid models of such networks are frequently not available for most distribution and transmission operators. Still, even when the grid model is at hand, the power system simulation carried out for the voltage profile estimation is computationally intensive, hindering real-time utilisation. Model-free voltage profile estimation analytics can thus respond to these challenges, but they require high-resolution net load data at each grid node as well as the corresponding electrical grid model for their assessment and validation.	
Primary Actor	The Distribution System Operator (DSO)
Secondary Actors	AI developers
AI techniques / Models	Usage of AI techniques: ML models such as Clustering
Equipment / Hardware used	Use of the RTPSS Laboratory
Software used	Python
Use Case Description	
Ensuring accurate load flow results of different conditions without using power flow calculation. Achieving low-latency processing for real-time response. Validating performance in realistic grid scenarios using CEDER-CIEMAT's Digital Twin.	
User Perspective	
The DSO is responsible for managing the distribution grid and requires a more streamlined solution to grid simulation beyond slow power flow calculations for grid impact calculations.	
System Perspective	
The model-free voltage estimation is designed to be modular and interoperable. It can be connected to grid planning and asset management platforms. Within the TEF, it interacts with the Digital Twin by receiving real voltage data as input and returning a simplified model of the grid. This loop allows the platform's behaviour to be studied and refined in controlled yet dynamic conditions, ensuring it can later be deployed in the real grid with minimal integration friction.	
Details	
Business Goals/Benefits	
The goal of this use case is to enable rapid and automated grid simulation and analysis without requiring the real and complete distribution grid model: • Faster evaluation of grid operation • Enhanced situational awareness for operators • Reduced manual workload and human error • Better grid planning and decision-making	
Application	The use case pertains to a model-free voltage estimation application. It processes voltage data

	and provides real-time voltage estimation based on grid changes.
Data Preconditions	The CEDER-CIEMAT microgrid data must be properly connected to output to the DT.
System Preconditions	Fully operational testing environment with access controls and user authentication.
User Preconditions	Users should be trained to interpret the output reports or interact with the analysis dashboard.
Trigger	AI developer completes prototype phase and requests real-world validation.
Challenges/Limitations	Cybersecurity and data privacy concerns during data transmission and processing.
Expected results - KPIs	
Accuracy: Phase and feeder identification accuracy is targeted to ~85%.	
User journeys	
- A user wants to simulate a change to a LV grid but does not have the grid model, only the grid's smart meters voltage data, and the knowledge of which smart meter will have a change (for example, increase/ decrease of contracted power or installation/ increase of PV generation). - The user submits the smart meter data and identifies the smart meter client adjustment through a .csv file. - The platform ingests and triggers an automated analysis. - Within seconds, a report is available to the user via .csv file with the estimated smart meter feeder and phase identification and the estimated max percentage change (up or down) in each clients' voltage profile, with an indication and timing of possible voltage alarms (i.e., under and overvoltage).	

3.2.2 Pilot 2 – CIEMAT – Microgrid-based Predictive Maintenance for Renewable Energy Equipment

Overview	
Owner of the Pilot	CIEMAT
Use Case Acronym	P2_UC1
Document Version	V1.0
Scope / Objectives of Use Case-Experiment	
Validate AI models for predictive maintenance of renewable energy equipment in real environments, using the CEDER-CIEMAT microgrid as a test bed.	

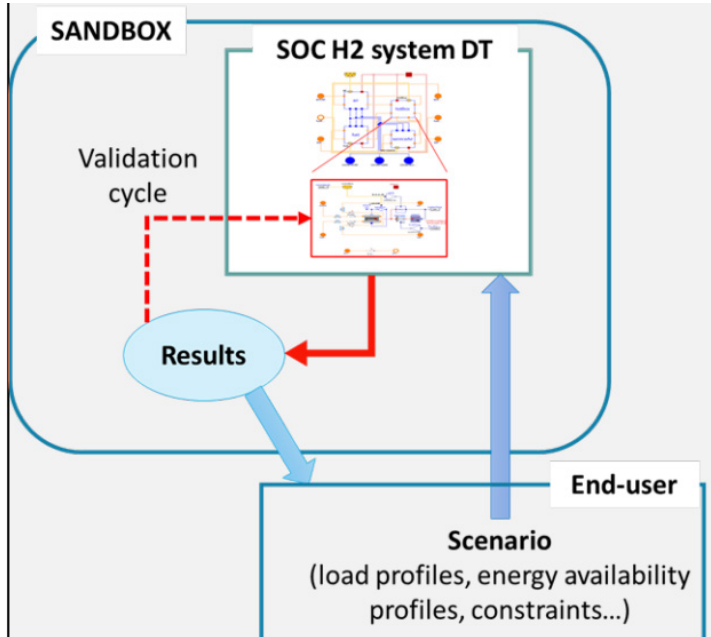
Current Situation	
CEDER-CIEMAT has an operational microgrid with multiple renewable sources (PV, wind, storage), monitored in real time. However, the data collected are not yet exploited with AI algorithms for predictive maintenance.	
Use Case	
Diagram of Use Case	
<pre> graph LR A["CEDER-CIEMAT Infrastructure • PV systems • Wind turbines • Storage (batteries) • EV charging points • SCADA / sensors / weather station"] --> B["Data Capture Layer • Real-time monitoring (1s/15min) • Local database / preprocessing scripts"] B --> C["Integration with EnergyGuard TEF • Data sharing (via secure API / export) • Digital twin synchronized with data • User access for model validation"] </pre>	
Primary Actor	Energy systems integrators / AI developers
Secondary Actors	Researchers, infrastructure managers, maintenance technicians.
AI techniques / Models	TBD
Equipment / Hardware used	Field sensors, database servers, weather stations, converters. Standard CPU. Data: up to 10MB/day, stored in local repositories.
Software used	Platforms based on Home Assistant, Node-RED.
Use Case Description	
Integration of predictive AI models that warn of impending failures or behavioural anomalies in equipment, enabling early preventive maintenance, reduced failures and improved performance.	
User Perspective	
It allows models to be validated in a real environment, with real data, reducing risks before deploying them at scale.	
System Perspective	
The existing system is connected to the TEF environment to share data and test algorithms under controlled and monitored conditions.	

Details	
Business Goals/Benefits	
Reduce downtime, reduce maintenance costs, increase reliability of distributed energy systems.	
Application	CEDER-CIEMAT hybrid micro-grid
Data Preconditions	Availability of historical data and real-time data (production, consumption, temperatures, etc.).
System Preconditions	Infrastructure and database operational; access to TEF platform enabled.
User Preconditions	Users with a technical profile and validated access to the platform.
Trigger	Need to validate a predictive failure/energy efficiency model.
Challenges/Limitations	Limited labelling of historical faults; dependence on quality of collected data; noise sensitive model.
Expected results - KPIs	
- Fault prediction accuracy >90%. - Reduced electrolyser downtime. - Improved operating efficiency (kWh/kg H₂). - Simulation of stress scenarios with validated AI.	
User Journeys	
- The user (AI developer or energy systems integrator) accesses the TEF platform portal after obtaining the corresponding credentials. - He/she explores the available datasets of the CEDER-CIEMAT microgrid, including historical data on renewable production, consumption and battery status. - Upload its failure prediction model to the platform or connect it via API to test it with real data from the microgrid. - Run tests on simulated scenarios, including events such as overloads, production variations and device outages. - Visualise the results obtained (failure prediction, model efficiency, alerts) and compare performance against expected metrics (KPIs). - Adjust your model based on feedback, re-validate with new data or more demanding scenarios. - Export the test results and obtain a report that validates the model for possible deployment in real installations.	

3.2.3 Pilot 3 – CEA, CARTIF, BER

This pilot consists of 3 different use cases, listed in the following sub-sections.

3.2.3.1 Use Case 1 – CEA – MURPHY Digital Twin

Overview	
Owner of the Pilot	CEA (Commissariat à l’Energie Atomique)
Use Case Acronym	P3_UC2
Document Version	V1.0
Scope / Objectives of Use Case-Experiment	
Validate a control law proposed by a user. Produce high fidelity results for a specified scenario for a SOC H2 production system.	
Current Situation	
The DT has already been used to test control laws before implementing them on the PLC and testing them in real situations.	
Use Case	
 The diagram illustrates the validation cycle within a sandbox environment. It features a 'SANDBOX' container holding the 'SOC H2 system DT'. A red dashed arrow labeled 'Validation cycle' points from the DT to a 'Results' oval. A blue arrow points from the 'Results' oval to an 'End-user' box. The 'End-user' box contains a 'Scenario' (load profiles, energy availability profiles, constraints...). A blue arrow points from the 'End-user' box back to the 'SOC H2 system DT'.	
The user will propose a scenario and will validate if the logic is ok or not. A fault occurs, how to manage it with an adjustment of the control law?	
Primary Actor	Industrials/engineer that want to test their control solutions. Researcher that wants to test their IA model using data, comparing results with digital twin.
Secondary Actors	The scenario creators. System control experts.
AI techniques / Mode	No AI techniques are in development.

Equipment / Hardware used	The hardware used is mainly composed of development computers with classic performances.
Software used	Digital twin: Simulink + Modelica
Use Case Description	
Testing the control law upstream will allow for optimization before going into production. In a system operational context, being able to simulate a fault and have a control solution without directly testing the solution in situ will be an advantage.	
User Perspective	
Users will have access to a high-fidelity DT, fed with real system data, for the development, the validation and the optimization of their control solution.	
Details	
Business Goals/Benefits	
The results obtained will help to design and specify their future electrolysis systems.	
Application	Hydrogen production
Data Preconditions	Have several scenarios available as well as simulation results
System Preconditions	Deployment of the digital twin
User Preconditions	Creation of a scenario to test
Trigger	A new scenario
Challenges/Limitations	Technical: usage of docker (trying to find a solution at CEA) To have a DT compatible with the diversity of scenario created.
Expected results - KPIs	
- Actual performance (e.g. execution time), known limitations, bottlenecks, etc. - Perspectives of evolution: Computing power, computing technology (CPU, GPU, FPGA), new algorithms, Application of AI, etc. - Quantification of needs: Speedup, increasing of data volume, improvement of accuracy, etc.	
User journeys	
The user connects to the TEF platform and provides its own scenario to the high temperature H2 production system Digital Twin. Then, if the inputs provided are validated automatically (compatible with the DT range of validity), the user can start the dynamic simulation and see the behavior of the chosen variables over time. The results are exportable for more detailed analysis offline.	

3.2.3.2 Use Case 2 – CARTIF – AI-Powered Digital Twin for Hybrid Energy System Optimization

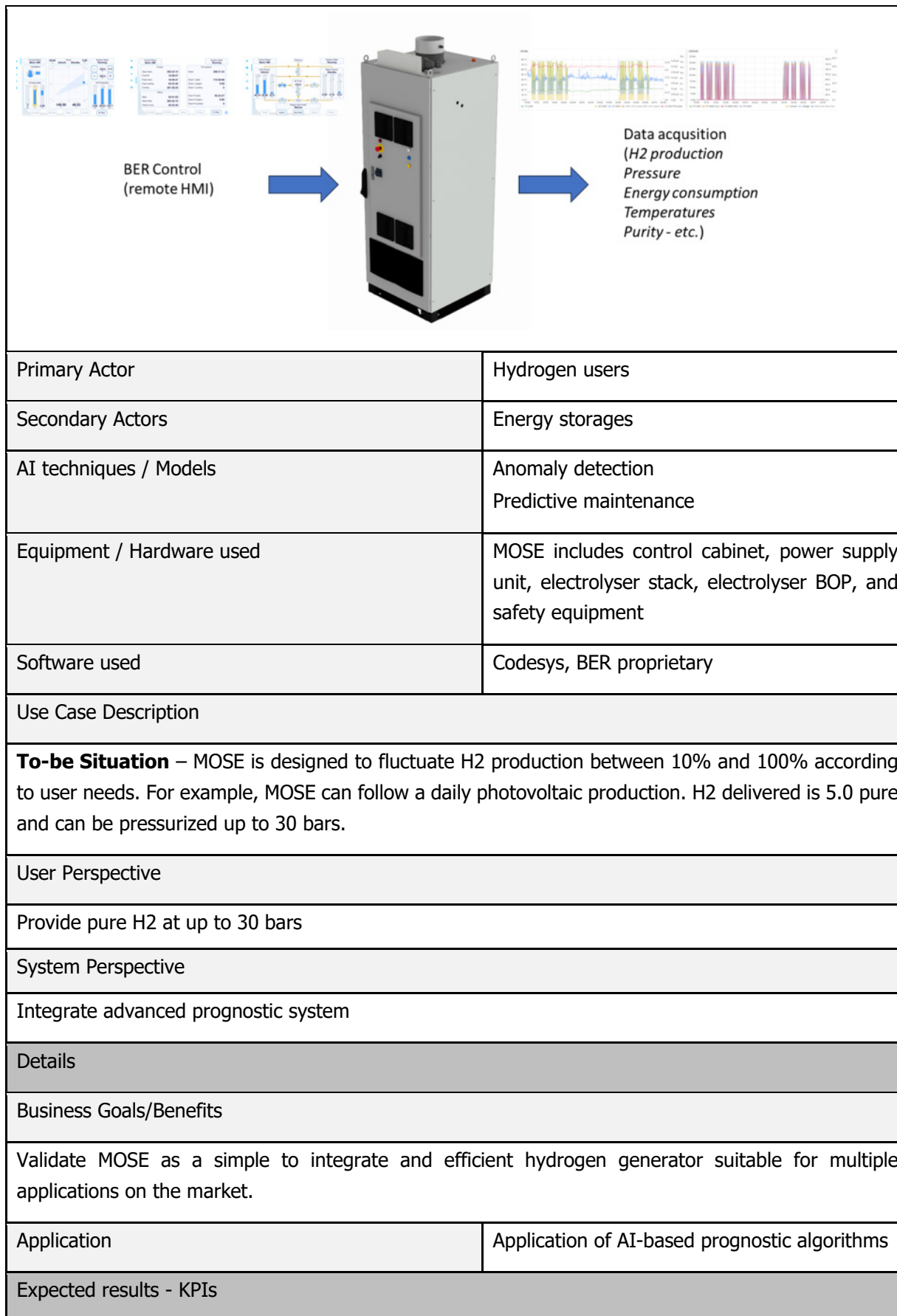
Overview	
Owner of the Pilot	CARTIF
Use Case Acronym	P3_UC2
Document Version	V1.0
Scope / Objectives of Use Case-Experiment	
Test the performance of the hybrid energy system configuration under the goal of optimizing the use of the resources, maximizing the renewable energy and fostering the use of H2. The system will operate under different conditions / parametrization to determine how to manage the energy resources according to the PV renewable available input.	
Current Situation	
As-is Situation – Data are being continuously monitored to extract operation and performance of the system. Moreover, initial models in Matlab and TRNSYS are made so that DTs can be developed following the control strategies.	
Use Case	
Primary Actor	Energy managers
Secondary Actors	AI developers, researchers
AI techniques / Models	Usage of AI techniques: - o ML/DL Models: Regression Analysis by using the Matlab and TRNSYS models, which, combined, will provide a grey-box approach to combine ML and physics to provide a more accurate model. - o Data sets: Energy flows, temperatures and loads ingested via MQTT broker. Data curation will be part of the data quality techniques within the data lake.
Equipment / Hardware used	Hardware / equipment used in the workflows: - o Computing technologies: CPU. - o Interconnection Fabric: Ethernet. - o Storage devices: Databases in local servers. - o 300MB/month to be stored and, in memory just MB to be managed (real-time data).
Software used	Software used in the workflows: o Maturity: Prototype stage with elements moving towards pilot validation and pre-industrial readiness.

	- o Sources: Combination of proprietary development (Matlab models, control logic) and open-source tools (TRNSYS, Python-based libraries for ML). - o Matlab/Simulink (proprietary, modelling and control logic) - o TRNSYS (commercial software, dynamic simulation) - o Python with libraries: Pandas, NumPy, Scikit-learn - o TensorFlow/PyTorch (open-source AI stack) - o MQTT for data ingestion (open source) - o OS: Windows and Linux (TRNSYS and Python environments) - o Languages: Matlab, Python
Use Case Description	
To-be Situation – AI-based DT with ingestion of real data and configuration / customization capabilities to manage the combination of multiple energy systems. This would allow running experiments and scenarios under different conditions to analyse both the operation strategies and the capabilities of the AI-based algorithms.	
User Perspective	
☐ Test the operation and performance of the hybrid energy system. ☐ Allow the interaction between the physical and virtual environments to run scenarios. ☐ Research about optimisation strategies over this configuration.	
Details	
Business Goals/Benefits	
To support decision-making in energy system design, improve operational efficiency, reduce emissions, and validate AI-based strategies for real-world deployment in residential and tertiary environments.	
Application	Application of AI-based digital twins to manage hybrid energy systems combining H2, batteries, and thermal storage for low-power use cases.
Data Preconditions	Availability of real-time energy flow, temperature, and load data via MQTT, curated and stored in a local data lake.
System Preconditions	Operational hybrid energy infrastructure with connected sensors, control units, and data acquisition.
User Preconditions	Access to platform interfaces, training on scenario execution, and permissions to interact with DT environments.
Trigger	A new scenario configuration, control strategy or external condition (e.g., change in PV input) initiates a test or simulation cycle.

Challenges/Limitations	High initial investment for real-world testing; complexity in modelling multi-vector systems; ensuring model accuracy; scalability of AI algorithms; data quality assurance.
Expected results - KPIs	
Energy self-consumption rate (%). Share of renewable energy used (%). Hydrogen utilisation efficiency. Prediction accuracy of the AI model (e.g. RMSE). Reduction in energy cost or carbon emissions. Time to convergence for control strategies or simulations.	
User Journeys	
The energy manager logs into the platform, selects a predefined scenario or provides a new configuration, and runs a simulation that adjusts system parameters. The AI-based DT provides feedback on energy balance, storage use, and emission reduction. Insights are used to refine the control strategy and support investment decisions.	

3.2.3.3 Use Case 3 – BER – DT-based MOSE Electrolyser Optimization

Overview	
Owner of the Pilot	BER
Use Case Acronym	P3_UC3
Document Version	V1.0
Scope / Objectives of Use Case-Experiment	
Optimize the operation of the Electrolyzer MOSE (MODular Smart Electrolyzer) - a product by BluEnergy Revolution. The tests that will be performed will allow the expand the library data collected on the product operation with the aim to optimize it and diffuse it on the market.	
Current Situation	
As-is Situation – The first MOSE electrolyzer units are reaching the first customers and applications; the system is currently being evaluated under real operative conditions. The data that will be provided by Energy Guard MOSE systems will be compared with the ones that are currently under acquisition.	
Use Case	
Diagram of Use Case	



Overall efficiency.
Identification of faults.
Predictive Maintenance.
Effective performance over time.

User journeys

TBD

3.2.4 Pilot 4 – REA – Digital Twin Assisted, AI-driven Decision Support for Enhanced Building Energy Efficiency

Overview	
Owner of the Pilot	REA
Use Case Acronym	P4_UC1
Document Version	V1.00
Scope / Objectives of Use Case-Experiment	
The objective is the validation and enhancement of an AI-driven decision support tool that aims to optimize renovation planning by providing users with data-driven insights, thereby promoting sustainable energy use and reducing costs in residential buildings. Integration of new data will enable ongoing improvement of the tool's predictive capabilities, ensuring its long-term effectiveness and usability.	
Current Situation	
Under development	
Use Case	
- AI-based decision support system for energy efficiency renovation actions. - Predicts energy savings and investment impact using real building data. - Suggests categorized renovation actions based on audit annex methodology - Fine-tuned via MLOps pipeline for better prediction accuracy AI I-ENERGY EFI-DSS  - Virtual replica of Riga's multi-apartment buildings. - Integrates real-time and historical building and environmental data. - Allows visualization of renovation scenarios and energy consumption impact. - Acts as testing and experimentation facility for the EFI-DSS tool. Riga Buildings Local Digital Twin  - Centralized storage for all structured and unstructured data sources. - Ingests data from REA, GEORIGA, SSCB's Construction Information System. - Stores anonymized energy consumption, audits, meteorological and heating data - Enables data curation, enrichment, and feeds to EFI-DSS for model training. Data Lake 	

Primary Actor	Residents
Secondary Actors	Municipality
AI techniques / Models	ML/DL models: physics-informed neural network model
Equipment / Hardware used	On-premises VMs/Servers
Software used	- OS/Programming Languages: Linux, Python (PyTorch), Javascript (React), PostgreSQL (PostGIS) - Sources: https://github.com/I-ENERGY/EEI_Service_frontend https://github.com/I-ENERGY/EEI_Service_backend - Parallelization: multi-core CPU environment.
Use Case Description	
The system is designed to provide an automated, data-driven approach for predicting building energy efficiency improvements along with their impact, based on minimal building-related input by the end-user. Future enhancements include incorporating large language models (LLMs) to improve the interpretability of AI recommendations, provide contextual insights, and generate user-friendly, tailored renovation advice.	
User Perspective	
The proposed physics-informed neural network model offers a significant leap forward for users by providing an automated, data-driven alternative to the current status quo of exhaustive, human-led energy audits and time-consuming simulations.	
System Perspective	
The TEF platform will ingest data from EnergyGuard’s data lake—including building information, audit data, and energy usage—into the EFI-DSS system to generate accurate predictions of building performance. In the future, GPU-accelerated large language models (LLMs) may be used to enhance analysis, generate insights, and enable natural language interaction. EnergyGuard users will benefit from smarter recommendations, clearer reporting, and easier integration with renovation planning and funding tools.	
Details	
Business Goals/Benefits	
- Supporting smart renovation decisions: EFI-DSS uses building data, audit results, and heating consumption to predict the impact of various renovation options, helping users choose the most effective and cost-efficient measures. - Enabling data-driven EPCs: The system estimates key Energy Performance Certificate components—like areas and U-values—to identify weak spots and guide targeted upgrades.	

- Improving accuracy with real-world data: Continuous ingestion of data from sources such as the Riga Energy Agency and the State Construction Control Bureau ensures model refinement and reliable outputs. - Promoting energy efficiency and equity: The pilot supports multi-apartment owners in planning renovations, ultimately increasing retrofit rates in Riga, reducing energy poverty, and strengthening trust in energy audits.	
Application	Application of AI-based supported decision-making for recommending energy efficiency improvements in buildings.
Data Preconditions	Energy audit data files must be processed and structured for accurate input into the tool. Other relevant datasets must be structured and categorized.
System Preconditions	Data integration of new real-world data into TEF platform and interaction between TEF platform, models, digital twin.
User Preconditions	Users are required to possess minimal input data – basic details about their buildings.
Trigger	The use case is triggered when a user wants to evaluate energy efficiency improvement options for a specific building.
Challenges/Limitations	Ensuring correct and high-quality preprocessing of available datasets remains a major hurdle. Existing data often requires significant cleaning, standardization, and transformation to be usable for the EFI-DSS use case. Inconsistent formats, missing values, and varying levels of detail make it difficult to prepare data that accurately reflects real-world conditions and supports reliable model predictions.
Expected results – KPIs	
• 1 software tool that provides renovation recommendations with energy and CO₂ savings, based on real data from Riga buildings. • 1 methodology for categorizing and digitizing energy audit actions, improving audit data quality and usability. • Decreased time by 90% for identifying existing energy performance of the building (current value = 2 weeks). • Digital Twins for ≥ 20 buildings in Riga. Increased awareness of the residents of ≥ 100 buildings in Riga on renovation benefits and investments, that will help to get a decision about renovation.	
• Prediction accuracy: $\geq 80\%$ class match on validation buildings with known EPCs ($\Delta\text{EPC} \leq 1$ class).	

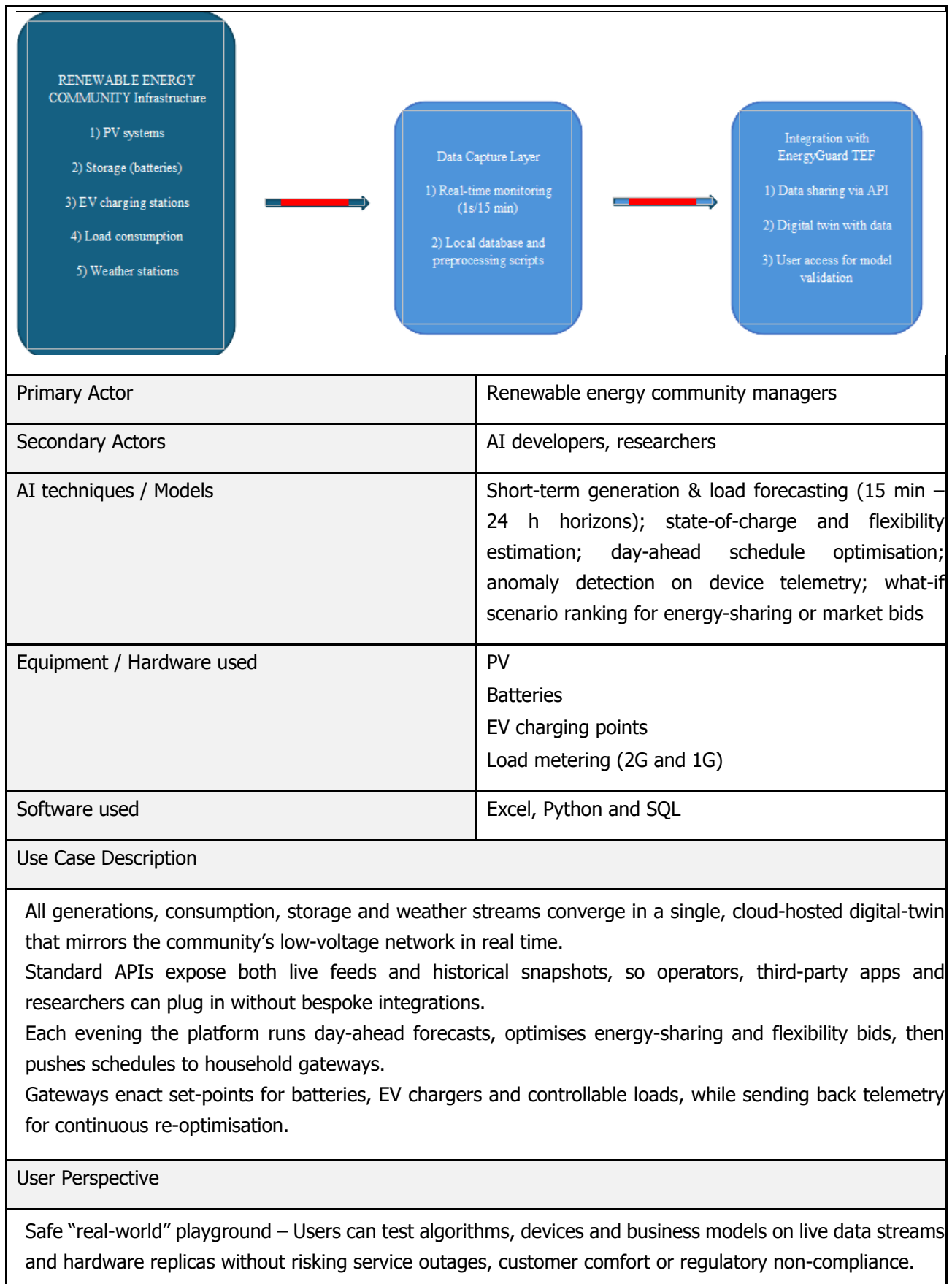
- Metadata completeness rate: $\geq 90\%$ of entries in the data lake shall include standardized metadata.

User journeys

Maria, a building administrator, logs into the EFI-DSS platform via EnergyGuard's secure portal. She selects her apartment complex using the building's address and basic characteristics. The system generates a virtual EPC based on available data and flags low insulation performance. Maria uses the simulation tool to explore renovation options—such as insulation upgrades—and reviews the projected EPC improvement and estimated savings (e.g. up to 22% reduction in energy demand). She downloads the simulation report containing cost estimates and energy impact and shares it with residents for discussion and approval of the renovation plan.

3.2.5 Pilot 5 – ENGREEN – Enhancing performance of Renewable Energy Community Facilities through AI experiments

Overview	
Owner of the Pilot	EnGreen
Use Case Acronym	P5_UC1
Document Version	V1.0
Scope / Objectives of Use Case-Experiment	
Pilot 5 will show how data-driven tools can strengthen cooperative energy management and unlock new value streams for local communities.	
Current Situation	
Planning and reporting are largely spreadsheet-based. There is no digital twin, no real-time forecasting, and no API through which third parties can test optimisation algorithms. Batteries, controllable loads and EV charging exist but have never been aggregated or offered to any local or national flexibility market. Generation (e.g., rooftop PV), consumption, storage and EV-charger measurements are collected, but remain in separate manufacturer portals with no unified access or common format. Data privacy, cyber-security and consent procedures differ across devices and operators, complicating external data sharing.	
Use Case	
Diagram of Use Case	



Faster, cheaper innovation cycle – Shared cloud/edge resources, standard APIs and ready-made datasets remove the upfront cost of building bespoke testbeds, so SMEs and researchers can focus budget and time on their core ideas. Objective performance validation – Common KPIs, reference scenarios and benchmarking tools let users prove the added value of their solutions against clear baselines, strengthening grant applications and investor pitches. Smoother path to market adoption – Results generated in the TEF carry recognised credibility with grid operators, regulators and customers, accelerating certification processes and scaling to other communities or markets.	
System Perspective	
Standards-based integration layer – each legacy EMS/SCADA or data portal connects to the TEF through lightweight adapters that translate native protocols (Modbus, MQTT, REST, IEC 61850, etc.) into a common event schema and API gateway. This lets the TEF ingest real-time telemetry and send back control set-points without altering the existing platforms' core logic.	
Details	
Business Goals/Benefits	
Demonstrate and quantify value – show, in a controlled yet realistic setting, how data-driven coordination of local generation, storage and flexible demand can cut energy costs, raise self-consumption and reduce CO₂, providing clear evidence for investors, grid operators and policymakers. Create a scalable reference model – produce open, standards-based methods, KPIs and business proofs that other communities and technology vendors can adopt quickly, shortening time-to-market and de-risking future deployments.	
Application	Cloud-edge DER orchestration platform integrated with the TEF sandbox: a digital-twin plus optimisation layer that balances real-time local supply and demand and aggregates bids for participation in emerging flexibility-services markets, while interfacing with existing EMS/SCADA through standards-based APIs.
Data Preconditions	Unified live feeds: All relevant assets (PV inverters, smart meters, batteries, EV chargers, weather station) must publish time-stamped measurements through a common, authenticated interface with clock drift < 15 minute and validated units. Clean historical dataset: At least 6-12 months of gap-filled, quality-checked time-series plus associated weather data must be available for baseline analysis and model training.

	Complete asset metadata: Each device requires harmonised records of rated power, location, phase/feeder, controllability flags and contractual attributes, stored in a single master-data schema. Legal & security readiness: User consent, data-sharing agreements and role-based access controls must be in place before ingestion and external experimentation can begin.
System Preconditions	Edge-to-cloud infrastructure in place: secure gateways at each asset site, reliable Internet / cellular backhaul ≥ 10 Mbps, and a cloud tenancy provisioned for the digital-twin, data-lake and AI services.
User Preconditions	Informed consent & onboarding completed: every participating resident, operator or external tester has signed the TEF data-sharing and remote-control agreement and has finished the privacy/cyber-security induction. Credentials & training in place: users have active TEF accounts with multi-factor authentication and have followed the short tutorial that explains dashboards, APIs and fallback procedures. Asset registration & preferences set: each owner has linked their devices to the platform, verified the asset metadata, and defined basic operating constraints (e.g., opt-out times, minimum battery state of charge), so optimisation can start without manual intervention.
Trigger	Activation signal from the TEF orchestrator: once all data, system and user preconditions are verified ("green status"), the TEF platform emits a Start-Experiment event that automatically spins up the digital-twin, forecasting and optimisation workflows for the pilot assets. Optional calendar / market gate alignment: the orchestrator typically schedules this signal for the first day of a predefined test window (e.g., at 00:00 UTC on the agreed go-live date) or immediately after a successfully cleared flexibility-market bid, ensuring that the experiment begins under stable operating conditions.

Challenges/Limitations	Data completeness & quality – gaps, time-drift or mis-labelled measurements can cripple forecasting and optimisation accuracy before the experiment even starts. Connectivity & device reliability – rural backhaul outages, gateway reboots or firmware mismatches may break real-time control loops and force fall-back operation. Cyber-security & privacy constraints – strict encryption, access control and GDPR requirements can slow onboarding of external partners and restrict model-sharing options. Regulatory and market uncertainty – evolving rules for community energy billing and flexibility trading may alter revenue streams mid-project, making business results hard to generalise. Stakeholder adoption & skills gap – limited technical skills or resistance to automated control among residents/operators can delay deployment and reduce the scale of usable assets.
Expected results - KPIs	
Key KPIs ☐ Forecast accuracy: Mean Absolute Error $\leq 5\%$ on 15-min PV + load predictions. ☐ Community self-consumption: +15 % versus today. ☐ Energy-bill saving: –10 % annual cost per member. Computing KPIs ☐ Daily optimisation runtime: < 10 min in cloud. ☐ Bottlenecks: cellular backhaul drops; small sample size for rare events. ☐ Algorithms: shift from CNN/LSTM to lightweight Transformers and federated learning for better generalisation under data-privacy constraints. ☐ AI applications: real-time anomaly detection, adaptive EV-charger scheduling, auto-tuning of market bidding strategies.	
User journeys	

User Journey 1: "Policy Maker" or "REC Manager" (Non-technical)

A public administrator or the manager of a Renewable Energy Community (REC), is guided through the setup of their REC by, defining the number and type of members (e.g. consumers, prosumers), uploading or manually entering historical data (production, consumption, etc.) and specifying basic configuration parameters (e.g. battery presence, sharing rules).

Once the REC is defined, the user sees intuitive dashboards showing self-consumption rates, estimated collective savings and emissions reduction and load coverage.

With one click, the user can simulate predefined or custom scenarios (e.g. adding a battery, changing sharing rules) to see the impact on KPIs such as bill savings and energy independence.

A concise report is generated automatically, summarizing the setup and results, ready to be shared with citizens or local stakeholders.

User Journey 2: "Developer" or AI Expert (Technical user)

A developer, researcher or startup receives API keys to interact with the testing environment. The user selects an existing REC configuration or creates a custom scenario by defining user profiles, asset types, topology and technical parameters such as battery size, thresholds, and flexibility limits.

The DT exposes a set of endpoints to retrieve simulated real-time or historical data (e.g. PV production, battery SOC) and query performance indicators such as self-consumption or economic benefit.

The developer runs their service and uses the API to retrieve inputs from the DT, compute control decisions (e.g. "charge the battery with 70% of excess PV") and send back actions to be executed in the simulated environment.

3.3 EnergyGuard TEF Specifications

This section presents the specifications for the EnergyGuard Testing and Experimentation Facility (TEF), derived from a structured analysis of stakeholder inputs and pilot use case requirements. The methodology used to define the TEF specifications follows a user-centred and iterative approach, combining qualitative data from the questionnaires, described in Section 3.1, with technical insights from the five pilot demonstrations, described in Section 3.2.

The process consisted of the following steps:

- **Stakeholder Segmentation**: Inputs from nine categories of stakeholders were collected through a combination of online questionnaires and direct interviews, to capture both broad trends and in-depth perspectives.
- **Thematic Synthesis**: Responses were clustered into common functional and non-functional themes, highlighting recurring concerns and expectations.
- **Persona Development**: Based on this synthesis, representative user personas were created to reflect typical user types across the TEF ecosystem. Two types of user personas were developed.
 - First, general personas were defined to represent archetypal users who interact with the EnergyGuard TEF as a platform—such as AI developers.
 - Second, pilot-specific personas were derived from the five real-world pilot applications, representing the final users of AI-based energy solutions deployed and validated

through the TEF. These pilot personas reflect the operational context and technical needs of the energy systems and services targeted in each use case.

- **User Story Mapping:** User stories were developed through interview/questionnaire inputs for both general and pilot-specific personas. The general user stories describe how different stakeholder types interact with the TEF platform and its core services, capturing typical workflows such as uploading AI models, running experiments, or validating compliance. In parallel, pilot-specific user stories reflect concrete use-case scenarios from the five pilots, describing how the TEF supports real-world applications of AI in the energy sector—such as optimizing grid operations, performing predictive maintenance, or improving energy efficiency.
- **Requirement Extraction:** Functional and technical requirements were derived from both general and pilot-specific user stories. The general requirements reflect the core functionalities expected from the TEF platform by its various user groups. The pilot-specific requirements, on the other hand, address the concrete technical and functional needs identified in the real-world use cases. All requirements were then categorised according to their priority, identifying the essential capabilities required to implement the Minimum Viable Product (MVP) of the TEF, ensuring it can support initial pilot deployments and basic user interactions.

3.3.1 User personas

User personas are representative profiles that capture the key characteristics, goals, needs, and behaviours of typical users who will interact with the EnergyGuard TEF and its applications. They are based on real data collected from stakeholders and pilots and serve to guide the design and development of user-centred functionalities and services. The following tables lists the user personas that have been identified. In particular, the first table lists personas that are involved in direct use of the TEF, whereas the second table lists personas that are involved in the use of pilot applications and other applications using the TEF.

ID	Name	Goals	Pain points
UP-TEF-001	AI Developer /startup	AI developers and startups want to use the TEF to refine and validate their AI applications. They need to validate models based on real data.	They do not have access to real environments with sufficient granularity, frequency and traceability of data.
UP-TEF-002	Researchers and Academia	Academic researchers want to develop AI models in different energy sectors, including microgrids and hybrid systems, and want to demonstrate performance and robustness of their models under real operating conditions.	They have difficulty in finding controlled environments that represent real energy systems.
UP-TEF-003	EPES stakeholders (Grid Operator / Utility Engineer,	They need to test AI algorithms in a realistic, large-scale environments, to ensure	Limited access to large-scale digital twin environments and privacy-compliant data.

	generators, suppliers, TSOs, DSOs, hydrogen players, aggregators, building operators, local energy communities, energy companies, suppliers, RECs, ESCOs, facility manager, RES plants O&M	stability and efficiency of energy operations / management using validated AI tools.	Lack of trustworthy AI solutions that comply with internal safety and regulatory policies.
UP-TEF-004	Administrator	Ensure the platform remains trustworthy and up to date by managing data, user roles, and localization settings.	Admins face challenges in keeping data current, ensuring security, and coordinating updates across diverse regional contexts and stakeholders.
UP-TEF-005	Policy Makers & Standardisation Organisations (SDO)	Promote the safe and effective deployment of AI in the energy sector by developing clear regulations and harmonized standards across borders.	Gaps in regulatory clarity, cybersecurity standards, and certification frameworks hinder both trust and innovation in AI-based energy solutions.
UP-TEF-006	Lawyers and Legal system entities	They need to ensure that AI experimentation in the energy sector complies with all relevant legal frameworks, while safeguarding consumer rights.	Lack of a dedicated, sector-specific regulatory sandbox makes it difficult to assess the legal implications of AI solutions within the complex energy regulatory landscape.

Table 3: User personas of the EnergyGuard TEF

ID	Name	Goals	Pain points
UP-APP-001	REC manager	They want to predict and understand REC potential for energy sharing and flexibility services.	They miss reliable historical data and express privacy concerns for individual members of the REC.
UP-APP-002	Citizen/resident	They need to get information about possible building energy efficiency, retrofits, select them and predict energy/CO2/financial savings.	The building of interest could possibly not be classified under a specific building series. This is an out-of-distribution sample for the AI algorithm potentially leading to inaccurate results.
UP-APP-003	Municipality	They need to get information about renovated/not renovated buildings in city, calculate required investments	Same as for UP-APP-002.

		for renovation and predict energy, CO2 savings.	
UP-APP-004	Energy Service Company (ESCO)	They need to test and validate the operation and performance of an energy system and get insights on citizens / household owner needs and improve their energy efficiency services.	The investments to test and validate the technology might be relevant.

Table 4: User personas of pilot applications**3.3.2 User stories**

User stories represent concrete use cases that describe how different users interact with the EnergyGuard TEF platform and its services. They were developed from identified user personas and enriched with contributions from the project's five pilots. User stories are useful for illustrating realistic workflows and highlighting users' operational and technical needs, serving as a basis for deriving the platform's functional and non-functional requirements. The stories presented here include both the main general scenarios, valid for the entire EnergyGuard ecosystem, and some specific stories related to individual pilot use cases. In particular, the first table lists user stories that involve direct use of the TEF, whereas the second table lists user stories that are involve the use of pilot applications and other applications using the TEF.

Story ID	Persona ID	User story
US-TEF- 001	UP-TEF-001	As an AI Developer, I want to compare my model's energy analytics, forecasting or maintenance algorithms using real-world data so that I can validate them and improve their reliability and accuracy.
US-TEF- 002	UP-TEF-001	As an AI developer, I want to predict energy performance of buildings depending on the selected energy efficiency measures, and to find optimal scenario of renovation savings per investments.
US-TEF- 003	UP-TEF-001	As an AI developer having created AI based energy analytics services, I want to use HPC to ensure faster training and parallel model tuning and ensemble training
US-TEF- 004	UP-TEF-001	As an AI developer, I want to evaluate the compliance and trustworthiness for my energy analytics AI based solution.
US-TEF- 005	UP-TEF-001	As an AI developer, I want to have easy access to cloud-based testbeds and digital twins to validate services and models.
US-TEF- 006	UP-TEF-001	As an AI developer, I want to have baseline data and models to validate against real use cases.
US-TEF- 007	UP-TEF-001	As an AI developer, I need safety, security and compliance scores for my solutions.

US-TEF- 008	UP-TEF-001	As an AI developer, I want to seamlessly access AI assets from European ecosystems such as AIOD, Data Spaces, DIHs, so that I can develop and test AI-based energy applications leveraging existing European infrastructure.
US-TEF- 009	UP-TEF-002	As an academic researcher, I want to simulate low-voltage grid behaviour under variable DER conditions, so that I can evaluate the effectiveness of model-free voltage estimation algorithms.
US-TEF- 010	UP-TEF-002	As an academic researcher, I want to test and demonstrate the performance of my microgrid and/or hybrid energy system optimisation models in real operating conditions so that I can validate their robustness and applicability.
US-TEF- 011	UP-TEF-002	As an academic researcher, I want to continually monitor models for bias, drift and reliability to improve trustworthiness and robustness.
US-TEF- 012	UP-TEF-001 UP-TEF-002	Academic researchers, data scientist, and SMEs intend to collaborate with energy experts and other AI teams to build more appealing and realistic solutions and potential co-development opportunities.
US-TEF- 013	UP-TEF-003	As an energy engineer, I want to use the digital twin to test my control law so that I can correct my solution and be sure that it works.
US-TEF- 014	UP-TEF-003	As an energy engineer, I want to find the best control solution to compensate for a fault or to achieve a given yield.
US-TEF- 015	UP-TEF-004	As an administrator, I want to be able to evaluate applicants in terms of alignment with project values.
US-TEF- 016	UP-TEF-004	As an administrator, I want to validate and update the source and timestamp of cost data for energy measures, so that users can trust the recommendations provided.
US-TEF- 017	UP-TEF-005	As a policy maker, I need anonymised insights about the status of the developed AI services over time to create better public policies and incentives.
US-TEF- 018	UP-TEF-006	As a lawyer, I want to know what are the most important legal/ethical risks that AI developers face at the AI solutions' development phase and be able to acknowledge their cause to appropriately adjust my legal consultation.

Table 5: User stories of the EnergyGuard TEF

Story ID	Persona ID	User story
US-P5C1-001	UP-APP-001	As a Renewable Energy Community (REC) manager, I want a plug-and-play digital-twin platform that gives me clear AI-based forecasts, automated battery/asset schedules and simple dashboards for members, so that I can maximise local renewable self-consumption, cut energy bills, and transparently share the benefits—even without deep technical expertise.
US-P4C1-002	UP-APP-002	As a citizen I am interested in energy efficiency level of my multiapartment building and economically viable measures for renovation. I want to calculate potential savings of energy, CO2 and euro.

US-P4C1-003	UP-APP-002	As a citizen, I want to consent explicitly before my building data is stored or shared, so I feel confident using the tool while retaining control over my information.
US-P4C1-004	UP-APP-003	As a municipality, I want to know how many buildings need to be renovated to define necessary investments and potential energy savings for the whole municipality.
US-P1C1-005	UP-APP-004	As an ESCO, I want to validate AI-based systems in a digital twin of our network, so that I can ensure grid stability before deploying them in the field.
US-P4C1-006	UP-APP-004	As an ESCO, I need access to virtual EPCs, alongside their selected measures and upgrades so that I can provide them with the necessary experts, services, and materials to implement them.
US-P3C2-007	UP-APP-004	As an ESCO interested in low-power and hybrid energy systems, I want to test and validate the operation and performance of integrated solutions combining batteries, hydrogen, and thermal storage, so that I can make informed investment decisions and ensure reliable, efficient deployment in real-world applications.
US-P1C1-008	UP-TEF-002	As a researcher, I want to simulate low-voltage grid behaviour under variable DER conditions, so that I can evaluate the effectiveness of model-free voltage estimation algorithms.
US-P1C1-009	UP-TEF-001	As a startup or AI developer, I want to validate my energy forecasting or maintenance algorithms using real-world microgrid data so that I can improve their reliability and accuracy.
US-P1C1-010	UP-TEF-002	As an academic researcher or engineer, I want to test and demonstrate the performance of my microgrid and/or hybrid energy system optimisation models in real operating conditions so that I can validate their robustness and applicability.
US-P1C1-011	UP-TEF-004	As an admin, I want to validate and update the source and timestamp of cost data for energy measures, so that citizens can trust the recommendations provided.
US-P1C1-012	UP-TEF-003	As an energy engineer, I want to use the digital twin to test my control law so that I can correct my solution and be sure that it works.
US-P1C1-013	UP-TEF-003	As an energy engineer, I want to find the best control solution to compensate for a fault.
US-P1C1-014	UP-TEF-003	As an energy engineer, I want to find the best control solution so that achieve a given yield.
US-P1C1-015	UP-TEF-001	As a researcher or AI developer I want to predict energy performance of building depending on the selected energy efficiency measures; to find optimal scenario of renovation savings per investments.

Table 6: User stories of pilot applications

3.3.3 Requirements

User requirements define the functional and technical capabilities that the EnergyGuard TEF platform must provide to effectively support its users and pilot implementations. These requirements were derived from both general and pilot-specific user stories, ensuring alignment with the practical needs of stakeholders and real-world scenarios. General requirements reflect the core functionalities expected across the entire TEF ecosystem, while pilot-specific requirements address targeted features and constraints relevant to individual pilot use cases. Additionally, legal requirements have been identified and included by the legal department of EnergyGuard. Said requirements are neither related to specific user personas nor user stories and relate to the general operation of the TEF under EU regulation. Note also that the next versions of this deliverable will further specify those requirements (or appropriately modify if considered necessary) and related functionalities aiming to aid the development of EnergyGuard platform.

To support prioritisation and effective implementation planning, all requirements have been categorised by their importance, highlighting those essential for delivering the Minimum Viable Product (MVP) by the end of the project. Additional non-MVP features will be considered for implementation based on feedback from users within or after the project duration based on their complexity, development effort and cost. This approach ensures that the TEF can support initial pilot deployments and basic user interactions from the outset.

The following tables present the identified requirements. In particular, the first table lists the general requirements applicable to the platform as a whole, whereas the second table outlines pilot-specific requirements that respond to the unique needs of each pilot context.

Req. ID	Functional / Non-functional	Related user story	Description	MVP Yes/No
UR-TEF-F-001	Functional	US-TEF-001, US-TEF-006	The platform should provide components for ML/DL models lifecycle management, model store and model versioning.	Yes
UR-TEF-F-002	Functional	US-TEF-005, US-TEF-013	The platform should interact with the five testing and experimentation facilities, providing pre-built testbeds simulating different energy scenarios.	Yes
UR-TEF-NF-001	Non-functional	US-TEF-003	The platform shall ensure uptime for testbed access to support continuous integration and testing.	No
UR-TEF-NF-002	Non-functional	US-TEF-018	Data should be anonymized and / or aggregated. All personal data shall be stored in encrypted form and managed in full alignment with GDPR.	Yes

UR-TEF-F-003	Functional	US-TEF-018	The system must allow secure access to all datasets and applications, with user authentication, permission control and data versioning. The system shall also request and record explicit consent from users before storing or processing any personal data.	Yes
UR-TEF-F-004	Functional	US-TEF-001, US-TEF-009	The system shall allow ingest, storage and visualisation of high frequency time series data, with export in standard formats (.csv, .xls).	Yes
UR-TEF-NF-003	Non-functional	US-TEF-003, US-TEF-004	The system shall ensure data traceability (access logs) and shall ensure data integrity through redundancy and recovery mechanisms.	Yes
UR-TEF-NF-004	Non-functional	US-TEF-004, US-TEF-007	The system shall expose trust-related metadata to the user (e.g., data source, model confidence, last update) to support informed decision-making and increase transparency results.	No
UR-TEF-NF-005	Non-functional	US-TEF-013	The system shall maintain compatibility with Digital Twins used by local actors (e.g., REA), ensuring data exchange, model updates, and simulation results are synchronized and traceable in the EnergyGuard data infrastructure.	No
UR-TEF-NF-006	Non-functional	US-TEF-014	The system shall provide results of the DT optimization algorithms within a limited time to assure user-friendly use.	No
UR-TEF-F-005	Functional	US-TEF-013	The system should be able to allow configuration parameters to customize the pilot-related DTs.	Yes
UR-TEF-F-006	Functional	US-TEF-009, US-TEF-010	The system should collect data from the multiple pilots, harmonise data samples and store datasets in the data lake, and then share data to be consumed by the DTs related to the pilots.	Yes
UR-TEF-NF-007	Non-functional	US-TEF-014	The system shall respect the limits of the physical systems to avoid damages when real control is performed over the energy assets.	No
UR-TEF-F-007	Functional	US-TEF-004, US-TEF-007	The system shall provide a compliance assessment dashboard showing alignment with key regulations (e.g., GDPR, AI Act, energy sector norms), together with automated trustworthiness evaluation tools (fairness, explainability, robustness) that should allow for exporting compliance reports for internal or investor use.	No

UR-TEF-F-008	Functional	US-TEF-018	The system shall provide access to ethical impact checklists and scoring templates.	No
UR-TEF-F-009	Functional	US-TEF-004	The system shall allow for easily connecting existing AI services with the TEF for compliance evaluation.	Yes
UR-TEF-F-010	Functional	US-TEF-006, US-TEF-008	The system shall provide a secure data sandbox with historical and real-time energy datasets, allowing API and interface support for integrating SME's AI solutions into the test environment.	Yes
UR-TEF-F-011	Functional	US-TEF-001, US-TEF-002	The system shall provide tools for A/B testing AI models on energy analytics KPIs (e.g., demand forecasting accuracy, anomaly detection), also providing an audit trail for validation runs and results, logs and model performance analytics.	No
UR-TEF-F-012	Functional	US-TEF-003	The system shall provide a self-service portal to request and schedule HPC resources, allowing access to GPU-enabled compute environments with containerization support, with features for usage monitoring and resource allocation quotas.	No
UR-TEF-F-013	Functional	US-TEF-003, US-TEF-010	The system shall provide support for large-scale model training and simulation pipelines.	No
UR-TEF-F-014	Functional	US-TEF-015	The applicant portal shall contain submission forms, asking for business and technical info, and provide a scoring engine for assessing alignment with TEF values (safety, sustainability, fairness), with automated flagging for missing or non-compliant submissions.	Yes
UR-TEF-F-015	Functional	US-TEF-015	The applicant portal shall provide a review workflow with approval/rejection tracking.	No
UR-TEF-F-016	Functional	US-TEF-006	The system shall provide access to labeled baseline datasets and reference models (e.g., open-source load forecasters) for benchmarking, and a leaderboard-style comparison interface for evaluating AI model.	Yes
UR-TEF-F-017	Functional	US-TEF-010	The system shall provide documentation and examples for real-world deployment constraints.	No
UR-TEF-	Functional	US-TEF-011	The system shall provide a continuous model monitoring pipeline with bias detection (e.g., demographic parity, disparate impact)	No

F-018			and drift detection on input features and predictions (statistical and concept drift).	
UR-TEF-F-019	Functional	US-TEF-011	The system shall provide a reliability scoring mechanism based on performance under different conditions (e.g., weather, grid load), and an alert system for detected performance degradation.	No
UR-TEF-F-020	Functional	US-TEF-012	The system shall provide a project collaboration space with shared repositories, forums, documentation hubs, co-development project templates, legal frameworks (e.g., IP sharing agreements) and a matching and networking tool for pairing SMEs, researchers, and domain experts.	Yes
UR-TEF-F-021	Functional	US-TEF-012	The system shall provide an event calendar for workshops, hackathons, and joint innovation challenges.	No
UR-TEF-F-022	Functional	US-TEF-007	The system shall provide an automated scoring system with configurable weights for safety (fail-safe behavior), security (vulnerabilities), and compliance, with color-coded report cards (e.g., green/yellow/red flags) and exportable certifications and audit logs for external stakeholders.	No
UR-TEF-F-023	Functional	US-TEF-007	The system shall contain a modular testing suite based on standards such as ISO/IEC 27001 for security, and NIST AI RMF.	No
UR-TEF-F-024	Functional	US-TEF-017	The system should provide an anonymized dashboard for policymakers to show aggregated analytics on AI services (use cases, performance metrics, compliance trends), with the possibility to filter by time, sector (e.g., buildings, grid, transport), and technology type.	No
UR-TEF-F-025	Functional	US-TEF-017	The system should provide a policy recommendation engine for policymakers highlighting gaps and trends on AI services.	No
UR-TEF-NF-008	Non-functional	US-TEF-018	The system must adopt secure data sharing protocols and access control to protect IP and privacy.	Yes
UR-TEF-F-026	Functional	US-TEF-018	The system should provide access to an AI risk database including legal and ethical risks to identify probable and most common risks in developing trustworthy and compliant AI solutions.	No
UR-TEF-	Non-functional	Legal. Not derived	The platform should be IP/Copyright compliant with regards to training, input and output data.	No

NF-009		from user story		
UR-TEF-NF-010	Non-functional	Legal. Not derived from user story	The platform should operate in compliance with AI Act if and to the extent applicable which needs to further be assessed. During the project, assessment for probable exceptions to AI Act's application with regards to AI systems developed, used or incorporated or any other operation will take place.	Yes
UR-TEF-NF-011	Non-functional	Legal. Not derived from user story	The Platform should be data protection/privacy compliant both at the development and operation phase (to the extent that processing of personal data takes place in each one of them) and will comply with data protection by design and data protection by default with regards to data lake creation, platform development, data processing, AI models, EnergyGuard website operation.	Yes
UR-TEF-NF-012	Non-functional	Legal. Not derived from user story	The platform should comply with the ethical requirements provided by "Ethics Guidelines for trustworthy AI" drafted by High level Expert Group on Artificial Intelligence (HLEG Guidelines) and by "Ethics by Design and Ethics of Use approaches for artificial intelligence" drafted by European Commission for the AI systems that will be developed and used to serve the purposes of the project.	No
UR-TEF-NF-013	Non-functional	Legal. Not derived from user story	The platform, being significant part of EnergyGuard scientific project, should operate in accordance with the European Code of Conduct for Research Integrity.	Yes

Table 7: Requirements for the EnergyGuard TEF

Req. ID	Functional / Non-functional	Related User Story	Description	MVP Yes/No
UR-APP-F-001	Functional	US-P1C1-005, US-P3C2-007	The pilot applications should allow the integration of data into the digital twins to test and validate AI algorithms.	Yes
UR-APP-F-002	Functional	US-P4C1-006	The system shall allow users to generate a virtual EPC based on minimal inputs like address, construction type, and materials. Each EPC prediction shall be tagged with the version of the prediction model used, ensuring traceability, reproducibility, and compliance with EnergyGuard audit policies.	No
UR-APP-F-003	Functional	US-P4C1-002	The system shall enable users to select from a list of renovation measures and view updated energy class, energy savings, and investment cost.	No

UR-APP-F-004	Functional	US-P4C1-004	The system shall support the generation of anonymized EPC trends and renovation scenarios across neighbourhoods for municipal planning.	No
UR-APP-F-005	Functional	US-P4C1-004	The system shall support integration with GeoRiga, allowing EPC and simulation results to be visualized on Riga's city GIS infrastructure.	No
UR-APP-F-006	Functional	US-P5C1-001	The system shall provide AI-based forecasting of energy production and consumption, tailored to each community member.	No
UR-APP-F-007	Functional	US-P5C1-001	The system shall include automated asset and battery scheduling to optimise self-consumption and cost savings.	No
UR-APP-F-008	Functional	US-P4C1-002	The system shall calculate potential energy, CO ₂ , and cost savings based on the selected renovation measures.	No
UR-APP-F-009	Functional	US-P4C1-003	The system shall request explicit user consent before collecting, storing, or sharing building data.	No
UR-APP-F-010	Functional	US-P3C2-007	The system shall support simulation and validation of integrated hybrid energy systems, including batteries, hydrogen, and thermal storage, under configurable operational scenarios.	No
UR-APP-F-011	Functional	US-P1C1-008	The platform allow interaction with the Digital Twin of the Portuguese Transmission Grid.	No
UR-APP-NF-001	Non-functional	US-P2C1-010	The platform shall ensure uptime for testbed access to support continuous integration and testing.	No
UR-APP-F-012	Functional	US-P2C1-010	The platform should provide components for ML/DL models lifecycle management, model store and model versioning.	No
UR-APP-NF-002	Non-functional	US-P4C1-016	Data should be anonymized and when not possible aggregated.	No
UR-APP-F-013	Functional	US-P2C1-011	The system must allow secure access to CIEMAT datasets (microgrid), with user authentication, permission control and data versioning.	No
UR-APP-F-014	Functional	US-P3C1-013	The system shall allow ingest, storage and visualisation of high frequency time series data (1s and 15min), with export in standard formats (.csv, .xls).	No
UR-APP-NF-003	Non-functional	US-P2C1-011	The system shall be compatible with common formats, shall ensure data traceability (access logs) and shall ensure data integrity through redundancy and recovery mechanisms.	No

UR-APP-F-015	Functional	US-P4C1-012	Admin users shall be able to add, remove, and update energy efficiency measures and their attributes (e.g., cost, description, thermal properties).	No
UR-APP-F-016	Functional	US-P4C1-016	The system shall request and record explicit consent from users before storing or processing any building-specific or personal data.	No
UR-APP-F-017	Functional	US-P4C1-012	The system shall enforce access restrictions based on user roles (e.g., admin, resident) using Keycloak integration.	No
UR-APP-NF-004	Non-functional	US-P4C1-016	All personal and building-specific data shall be stored in encrypted form and managed in full alignment with GDPR.	No
UR-APP-NF-005	Non-functional	US-P4C1-012	The system shall expose trust-related metadata to the user (e.g., data source, model confidence, last update) to support informed decision-making and increase transparency of EPC results.	No
UR-APP-NF-006	Non-functional	US-P4C1-016	Each EPC prediction shall be tagged with the version of the prediction model used, ensuring traceability, reproducibility, and compliance with EnergyGuard audit policies.	No
UR-APP-NF-007	Non-functional	US-P1C1-008	The system shall maintain compatibility with building Digital Twins used by local actors (e.g., REA), ensuring data exchange, model updates, and simulation results are synchronized and traceable in the EnergyGuard data infrastructure.	No
UR-APP-F-018	Functional	US-P3C1-015	The platform must allow users to enter new scenarios.	No
UR-APP-NF-008	Non-functional	US-P3C1-014	The system shall provide results of the DT optimization algorithms within a limited time to assure user-friendly use.	No
UR-APP-F-019	Functional	US-P3C1-014	The system should be able to allow configuration parameters to customize the pilot-related DTs.	No
UR-APP-F-020	Functional	US-P2C1-011	The system should collect data from the multiple pilots, harmonise data samples and store datasets in the data lake.	No
UR-APP-F-021	Functional	US-P1C1-008	The system should share data to be consumed by the DTs related to the pilots.	No
UR-APP-NF-009	Non-functional	US-P3C1-013	The system shall respect the limits of the physical systems to avoid damages when real control is performed over the energy asset.	No

Table 8: Requirements for pilot applications

Annex I: Bibliography

- [1] Time, How AI Is Making Buildings More Energy-Efficient. Available online: <https://time.com/7201501/ai-buildings-energy-efficiency/>.
- [2] Smart Building International, From individual buildings to smart cities: digital twins, ai and IoT set to drive sustainability, new research reveals. Available online: <https://smartbuildingmag.com/application-stories/90524-from-individual-buildings-to-smart-cities-digital-twins%2C-ai-and-iot-set-to-drive-sustainability%2C-new-research-reveals>.
- [3] Sghiri, A., Gallab, M., Merzouk, S., & Assoul, S. (2025). Leveraging Digital Twins for Enhancing Building Energy Efficiency: A Literature Review of Applications, Technologies, and Challenges. *Buildings*, 15(3), 498. <https://doi.org/10.3390/buildings15030498>.
- [4] Yitmen, I., Almusaed, A., Hussein, M., & Almssad, A. (2025). AI-Driven Digital Twins for Enhancing Indoor Environmental Quality and Energy Efficiency in Smart Building Systems. *Buildings*, 15(7), 1030. <https://doi.org/10.3390/buildings15071030>.
- [5] Arsecularatne, B., Rodrigo, N., & Chang, R. (2024). Digital Twins for Reducing Energy Consumption in Buildings: A Review. *Sustainability*, 16(21), 9275. <https://doi.org/10.3390/su16219275>
- [6] Zhansheng Liu, Mingming Li, Weiyu Ji, Development and application of a digital twin model for Net zero energy building operation and maintenance utilizing BIM-IoT integration, *Energy and Buildings*, Volume 328, 2025, 115170, ISSN 0378-7788, <https://doi.org/10.1016/j.enbuild.2024.115170>.
- [7] Hui Liu, Zhe Du, Tingting Xue, Tao Jiang, An integrated artificial intelligence-driven approach to multi-criteria optimization of building energy efficiency and occupants' comfort: A case study, *Journal of Building Engineering*, Volume 103, 2025, 111944, ISSN 2352-7102, <https://doi.org/10.1016/j.jobbe.2025.111944>.
- [8] Mohammad H. Mehraban, Samad ME Sepasgozar, Alireza Ghomimoghdam, Behrouz Zafari, AI-Enhanced Automation of Building Energy Optimization Using a Hybrid Stacked Model and Genetic Algorithms: Experiments with Seven Machine Learning Techniques and a Deep Neural Network, *Results in Engineering*, 2025, 104994, ISSN 2590-1230, <https://doi.org/10.1016/j.rineng.2025.104994>.
- [9] Yang, X., Liu, Z., Tang, X., Lu, R., & Liu, B. (2023). An Efficient and Multi-private Key Secure Aggregation for Federated Learning. *arXiv preprint arXiv:2306.08970*. <https://arxiv.org/abs/2306.08970>.
- [10] Bonawitz, K., Ivanov, V., Kreuter, B., et al. (2017). Practical Secure Aggregation for Privacy-Preserving Machine Learning. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS '17)*.
- [11] Aziz, R., Banerjee, S., & Bouzefrane, S. (2024). Privacy Preserving Federated Learning: A Novel Approach for Combining Differential Privacy and Homomorphic Encryption. *Information Security Theory and Practice*, Springer. https://doi.org/10.1007/978-3-031-60391-4_11.
- [12] Jin, W., Yao, Y., Han, S., Gu, J., Joe-Wong, C., Ravi, S., Avestimehr, S., & He, C. (2023). FedML-HE: An Efficient Homomorphic-Encryption-Based Privacy-Preserving Federated Learning System. *arXiv preprint arXiv:2303.10837*. <https://arxiv.org/abs/2303.10837>.

-
- [13] Zhou, Y., Wu, Y., Pan, S., & Zhou, M. (2022). Federated Deep Reinforcement Learning for Energy Management in Smart Microgrids. arXiv preprint arXiv:2211.03430. <https://arxiv.org/abs/2211.03430>
- [14] Lyer, Siddharth et al. (2024-06). "Review of experimental and modelling investigations for solid oxide electrolysis technology". International Journal of Hydrogen Energy 72, pp. 537–558. DOI: 10.1016/j.ijhydene.2024.05.312.
- [15] Majumer S, et al., "Exploring the capabilities and limitations of large language models in the electric energy sector", 2024
- [16] Ruan J., et al., "Applying Large Language Models to Power Systems: Potential Security Threats", 2024
- [17] Huang C., et al., "Large Foundation Models for Power Systems", 2024
- [18] González-Briones, Alfonso, et al. "Energy optimization using a case-based reasoning strategy." Sensors 18.3 (2018): 865.
- [19] S. M. Lundberg, S. I. Lee, A unified approach to interpreting model predictions, in: Advances in Neural Information Processing Systems, Vol. 2017-December, 2017
- [20] M. T. Ribeiro, S. Singh, C. Guestrin, "Why Should I Trust You?": Explaining the Predictions of Any Classifier, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, Association for Computing Machinery, New York, NY, USA, 2016, p. 1135–1144.
- [21] C. Molnar, Interpretable Machine Learning, 2nd Edition, 2022
- [22] D. W. Apley, J. Zhu, Visualizing the effects of predictor variables in black box supervised learning models, Journal of the Royal Statistical Society Series B: Statistical Methodology 82 (4) (2020) 1059–1086
- [23] M. T. Keane, B. Smyth, Good counterfactuals and where to find them: A case-based technique for generating counterfactuals for explainable AI (XAI), in: I. Watson, R. Weber (Eds.), Case-Based Reasoning Research and Development, Springer International Publishing, Cham, 2020
- [24] A. Shrikumar, P. Greenside, A. Kundaje, Learning important features through propagating activation differences (2019) arXiv:1704.02685
- [25] M. T. Ribeiro, S. Singh, C. Guestrin, Anchors: High-precision model-agnostic explanations, Proceedings of the AAAI Conference on Artificial Intelligence 32 (1) (Apr. 2018)
- [26] H. Nori, S. Jenkins, P. Koch, R. Caruana, InterpretML: A Unified Framework for Machine Learning Interpretability, arXiv preprint arXiv:1909.09223 (2019)
- [27] V. Arya, R. K. E. Bellamy, P.-Y. Chen, A. Dhurandhar, M. Hind, S. C. Hoffman, S. Houde, Q. V. Liao, R. Luss, A. Mojsilovic, S. Mourad, P. Pedemonte, R. Raghavendra, J. Richards, P. Sattigeri, K. Shanmugam, M. Singh, K. R. Varshney, D. Wei, Y. Zhang, One Explanation Does Not Fit All: A Toolkit and Taxonomy of AI Explainability Techniques (2019).
- [28] R. K. E. Bellamy, K. Dey, M. Hind, S. C. Hoffman, S. Houde, K. Kannan, P. Lohia, J. Martino, S. Mehta, A. Mojsilovic, S. Nagar, K. N. Ramamurthy, J. Richards, D. Saha, P. Sattigeri, M. Singh, K. R. Varshney, Y. Zhang, AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias (Oct. 2018)
-

-
- [29] J. Klaise, A. V. Looveren, G. Vacanti, A. Coca, Alibi Explain: Algorithms for Explaining Machine Learning Models, *Journal of Machine Learning Research* 22 (181) (2021)
- [30] Volk, R., Stengel, J., & Schultmann, F. (2014). Building Information Modeling (BIM) for existing buildings — Literature review and future needs. *Automation in Construction*, 38, 109–127
- [31] Boje, C., Guerriero, A., Kubicki, S., & Rezgui, Y. (2020). Towards a semantic construction digital twin: Directions for future research. *Automation in Construction*, 114, 103179
- [32] Zhang, Y., Wang, W., Wang, Z., Gao, M., Zhu, L., & Song, J. (2021). Green building design based on solar energy utilization: Take a kindergarten competition design as an example. *Energy Reports*, 7, 1297–1307
- [33] Lyer, Siddharth et al. (2024-06). “Review of experimental and modelling investigations for solid oxide electrolysis technology”. *International Journal of Hydrogen Energy* 72, pp. 537–558
- [34] Z. Li, H. Zhang, H. Xu, and J. Xuan, “Advancing the multiscale understanding on solid oxide electrolysis cells via modelling approaches: A review,” *Renew. Sustain. Energy Rev.*, vol. 141, p. 110863, May 2021
- [35] M. Erdem Günay, N. Alper Tapan, Analysis of PEM and AEM electrolysis by neural network pattern recognition, association rule mining and LIME, *Energy and AI*, Volume 13, 2023, 100254, ISSN 2666-5468
- [36] Rui Ding, Yawen Chen, Zhiyan Rui, Kang Hua, Yongkang Wu, Xiaoke Li, Xiao Duan, Jia Li, Xuebin Wang, Jianguo Liu, Machine learning utilized for the development of proton exchange membrane electrolyzers, *Journal of Power Sources*, Volume 556, 2023, 232389, ISSN 0378-7753
- [37] G. Botta, M. Romeo, A. Fernandes, S. Trabucchi, P.V. Aravind, Dynamic modeling of reversible solid oxide cell stack and control strategy development, *Energy Conversion and Management*, Volume 185, 2019
- [38] Antti Pohjoranta, Matias Halinen, Jari Pennanen, Jari Kiviaho, Model predictive control of the solid oxide fuel cell stack temperature with models based on experimental data, *Journal of Power Sources*, Volume 277, 2015
- [39] Y. Chi, J. Lin, P. Li, and Y. Song, “Investigating the performance of a solid oxide electrolyzer multi-stack module with a multiphysics homogenized model,” *J. Power Sources*, vol. 594, p. 234019, Feb. 2024
- [40] C. Yang et al., “Modeling and analysis of the multiphysics transport parameters of a kilowatt-class multistack module of reversible solid oxide cells,” *Appl. Therm. Eng.*, vol. 235, p. 121373, Nov. 2023
- [41] N. V. Giridhar, D. A. Allan, M. Li, S. E. Zitney, L. T. Biegler, and D. Bhattacharyya, “Optimal operation of solid-oxide electrolysis cells considering long-term chemical degradation,” *Energy Convers. Manag.*, vol. 319, p. 118950, Nov. 2024
- [42] Ahmed, S. F., Alam, M. S. B., Hassan, M., Rozbu, M. R., Ishtiak, T., Rafa, N., ... & Gandomi, A. H. (2023). Deep learning modelling techniques: current progress, applications, advantages, and challenges. *Artificial Intelligence Review*, 56(11), 13521-13617
- [43] Shiri, F. M., Perumal, T., Mustapha, N., & Mohamed, R. (2023). A comprehensive overview and comparative analysis on deep learning models: CNN, RNN, LSTM, GRU. *arXiv preprint arXiv:2305.17473*
- [44] Papastefanopoulos, V., Linardatos, P., Panagiotakopoulos, T., & Kotsiantis, S. (2023). Multivariate time-series forecasting: A review of deep learning methods in internet of things applications to smart cities. *Smart Cities*, 6(5), 2519-2552
-

-
- [45] Najafzadeh, M., Pouladi, J., Daghigh, A., Beiza, J., & Abedinzade, T. (2024). Fault detection, classification and localization along the power grid line using optimized machine learning algorithms. *International Journal of Computational Intelligence Systems*, 17(1), 49
- [46] Javaid, N., Qasim, U., Yahaya, A. S., Alkhamash, E. H., & Hadjouni, M. (2022). Non-technical losses detection using autoencoder and bidirectional gated recurrent unit to secure smart grids. *IEEE Access*, 10, 56863-56875
- [47] Kiasari, M., Ghaffari, M., & Aly, H. H. (2024). A Comprehensive Review of the Current Status of Smart Grid Technologies for Renewable Energies Integration and Future Trends: The Role of Machine Learning and Energy Storage Systems. *Energies*, 17(16), 4128
- [48] Balamurugan, M., Narayanan, K., Raghu, N., Arjun Kumar, G. B., & Trupti, V. N. (2025). Role of artificial intelligence in smart grid - a mini review. *Frontiers in artificial intelligence*, 8, 1551661
- [49] Guato Burgos, M. F., Morato, J., & Vizcaino Imacaña, F. P. (2024). A Review of Smart Grid Anomaly Detection Approaches Pertaining to Artificial Intelligence. *Applied Sciences*, 14(3), 1
- [50] Wen, X., Liao, J., Niu, Q. et al. Deep learning-driven hybrid model for short-term load forecasting and smart grid information management. *Sci Rep* 14, 13720 (2024)
- [51] Alhanaf, A. S., Balik, H. H., & Farsadi, M. (2023). Intelligent Fault Detection and Classification Schemes for Smart Grids Based on Deep Neural Networks. *Energies*, 16(22), 7680
- [52] Fayyaz, S., Qureshi, R. S., Rehman, A. U., & Murtaza, M. A. (2024). A review of machine learning and AI-based approaches for building energy management. *Energy Reports*, 10, 161–180
- [53] Bouchiha, H., Bensaid, M., & Fouchal, H. (2021). Deep learning-based meta-modeling and optimization for building energy and comfort performance. *arXiv*
- [54] González-Burgos, C. A., & Rojo, J. C. (2024). Artificial Intelligence in sustainable building renovation: A systematic review. *Sustainability*, 16(9), 3627
- [55] Arévalo, P., Ochoa-Correa, D., & Villa-Ávila, E. (2024). Optimizing microgrid operation: Integration of emerging technologies and artificial intelligence for energy efficiency. *Electronics*, 13(18), 3754
- [56] Hernández, J. L., de Miguel, I., Vélez, F., & Vasallo, A. (2024). Challenges and opportunities in European smart buildings energy management: A critical review. *Renewable and Sustainable Energy Reviews*, 199, 114472
- [57] Nebey, A. H. (2024). Recent advancement in demand side energy management system for optimal energy utilization. *Energy Reports*, 11, 5422-5435
- [58] Köse, A., Sukhanov, I., Belikov, J., Tepljakov, A., & Petlenkov, E. (2024, June). Demand Side Management for Commercial Buildings: Multiple Heating Systems Based Optimization in Real-Time. In *2024 IEEE International Conference on Engineering, Technology, and Innovation (ICE/ITMC)* (pp. 1-8). IEEE
- [59] Khan, M. A., Saleh, A. M., Waseem, M., & Sajjad, I. A. (2022). Artificial intelligence enabled demand response: Prospects and challenges in smart grid environment. *IEEE Access*, 11, 1477-1505
- [60] Zatti, M., Moncecchi, M., Gabba, M., Chiesa, A., Bovera, F., & Merlo, M. (2021). Energy Communities Design Optimization in the Italian Framework. *Applied Sciences*, 11(11), 5218
- [61] P. R. Mazzarino, M. Capone, E. Guelpa, V. Verda, L. Bottaccioli and E. Patti, "A Multi-Agent framework to evaluate energy flexibility in District Heating networks," 2022 IEEE International Conference on Environment

- and Electrical Engineering and 2022 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe), Prague, Czech Republic, 2022, pp. 1-6, doi: 10.1109/EEEIC/ICPSEurope54979.2022.9854622
- [62] Twaisan, K., & Barışçi, N. (2022). Integrated distributed energy resources (DER) and microgrids: Modeling and optimization of DERs. *Electronics*, 11(18), 2816
- [63] Li, K., Zhang, T., & Liu, H. (2020). Deep reinforcement learning for multi-objective energy management in hybrid systems. *Applied Energy*, 275, 115398. <https://doi.org/10.1016/j.apenergy.2020.115398>
- [64] Zhang, Y., Sun, M., & Zhou, Q. (2021). Multi-objective optimisation for smart energy systems: A comprehensive review. *Renewable and Sustainable Energy Reviews*, 135, 110206. <https://doi.org/10.1016/j.rser.2020.110206>
- [65] Saleem, M. U., Shakir, M., Usman, M. R., Bajwa, M. H. T., Shabbir, N., Shams Ghahfarokhi, P., & Daniel, K. (2023). Integrating smart energy management system with internet of things and cloud computing for efficient demand side management in smart grids. *Energies*, 16(12), 4835.
- [66] G. A. Susto, A. Schirru, S. Pampuri, S. McLoone, and A. Beghi, "Machine Learning for Predictive Maintenance: A Multiple Classifier Approach," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 3, pp. 812–820, June 2015
- [67] Ucar, A.; Karakose, M.; Kırımça, N. Artificial Intelligence for Predictive Maintenance Applications: Key Components, Trustworthiness, and Future Trends. *Appl. Sci.* 2024, 14, 898
- [68] Onanena, L Oukhellou, E Côme, D Candusso, D Hissel, P Aknin, Fault-diagnosis of PEM fuel cells using electrochemical spectroscopy impedance, *IFAC Proceedings Volumes*, Volume 45, Issue 21. <https://doi.org/10.3182/20120902-4-FR-2032.00114>
- [69] D. Hissel, M.C. Pera, Diagnostic & health management of fuel cell systems: Issues and solutions, *Annual Reviews in Control*, Volume 42, 2016
- [70] Pierpaolo Polverino, Marco Sorrentino, Cesare Pianese, A model-based diagnostic technique to enhance faults isolability in Solid Oxide Fuel Cell systems, *Applied Energy*, Volume 204, 2017
- [71] Pahon, E., Yousfi-Steiner, N., Jemei, S., Hissel, D., & Moçoteguy, P. (2017). A Non-Intrusive Signal-Based Method for a Proton Exchange Membrane Fuel Cell Fault Diagnosis. *Fuel Cells*, 17(2), 238-246
- [72] Fouquet, N., Doulet, C., Nouillant, C. J. J., Dauphin-Tanguy, G., & Ould-Bouamama, B. (2006). Model based PEM fuel cell state-of-health monitoring via ac impedance measurements. *Journal of Power Sources*, 159(2), 905-913
- [73] Ghorbani, M., Boley, M., Nakashima, P. N. H., & Birbilis, N. (2024). An active machine learning approach for optimal design of magnesium alloys using Bayesian optimisation. *Scientific Reports*, 14(1), 8299
- [74] Bressel, M., Hilairet, M., Hissel, D., & Bouamama, B. O. (2016). Extended Kalman filter for prognostic of proton exchange membrane fuel cell. *Applied Energy*, 164, 220-227
- [75] Saha B, Goebel K, Christophersen J. Comparison of prognostic algorithms for estimating remaining useful life of batteries. *Transactions of the Institute of Measurement and Control*. 2009;31(3-4):293-308
- [76] Nitsche, C., Schroedl, S., & Weiss, W. (2004, June). Onboard diagnostics concept for fuel cell vehicles using adaptive modelling. In *IEEE Intelligent Vehicles Symposium*, 2004 (pp. 127-132). IEEE

- [77] Jemei, S. (2004). Neural Network Model of a PEM Fuel Cell Power Generator (No. FRNC-TH--11995). Universite de Technologie de Belfort Montbeliard et Universite de Franche Comte, Laboratoire de recherche en Electronique Electrotechnique et Systemes-L2ES, Unite Mixte UTBM-UFC-Inrets (France).
- [78] Subotić, V., & Hochenauer, C. (2022). Analysis of solid oxide fuel and electrolysis cells operated in a real-system environment: State-of-the-health diagnostic, failure modes, degradation mitigation and performance regeneration. *Progress in Energy and Combustion Science*, 93, 101011
- [79] Ghorbani, M., Boley, M., Nakashima, P. N. H., & Birbilis, N. (2024). An active machine learning approach for optimal design of magnesium alloys using Bayesian optimisation. *Scientific Reports*, 14(1), 8299. <https://doi.org/10.1038/s41598-024-59100-9>
- [80] Adversarial Attacks and Defenses in Deep Learning" - Kiu Ren, Thianhang Zheng, Zan Quin, Xue Liu, 2020
- [81] A Comprehensive Review of Adversarial Attacks in Machine Learning - Abdirashid Abukar Ahmed, Dr. Nirvair Neeru, 2024
- [82] Bridging Machine Learning and Cryptography in Defence Against Adversarial Attacks - Olga Taran, Shideh Rezaeifar & Slava VoloShinovskiy, 2019
- [83] Stateful Detection of Black-Box Adversarial Attacks" - Steven Chen, Nicholas Carlini, D. Wagner, 2020
- [84] Causal Effects of Adversarial Attacks on AI Models in 6G Consumer Electronics" - Da Guo, Zhengjie Feng, Zhen Zhang, Fazlullah Khan, Chien-Ming Chen, Ruibin Bai, 2024
- [85] Study on Adversarial Attacks Techniques, Learning Methods and Countermeasures: Application to Anomaly Detection" - Anis Bouaziz, Manh-Dung Nguyen, Valeria Valdés, Ana Cavalli, Wissam Mallouli, 2023
- [86] Trustworthy AI Principles to Face Adversarial Machine Learning: A Novel Study" - CH.E.N. Sai Priya , Manas Kumar Yogi, 2023
- [87] Chien-fei Chen, Rebecca Napolitano, Yuqing Hu, Bandana Kar, Bing Yao, Addressing machine learning bias to foster energy justice, *Energy Research & Social Science*, Volume 116, 2024, 103653, ISSN 2214-6296
- [88] Md Morshed Alam et al, Artificial intelligence integrated grid systems: Technologies, potential frameworks, challenges and research directions, *renewable and sustainable Energy Reviews* 211 (2025) 115251
- [89] Daut, M. A. M., Hassan, M. Y., Abdullah, H., Rahman, H. A., Abdullah, M. P., & Hussin, F. (2017). Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review. *Renewable and Sustainable Energy Reviews*, 70, 1108-1118
- [90] Leslie, D., Rincón, C., Briggs, M., Perini, A., Jayadeva, S., Borda, A., Bennett, S.J. Burr, C., Aitken, M., Katell, M., Fischer, C. (2023). *AI Ethics and Governance in Practice: An Introduction*. The Alan Turing Institute
- [91] Big-data for building energy performance: Lessons from assembling a very large national database of building energy use Paul A. Mathew et al, *Applied energy* 140 (2015) 85-93
- [92] Leslie, D., Burr, C., Aitken, M., Cows, J., Katell, M., and Briggs, M. (2021). *Artificial intelligence, human rights, democracy, and the rule of law: a primer*. The Council of Europe

-
- [93] Leslie, D. (2019). Understanding artificial intelligence ethics and : A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute
- [94] Meiser, M.; Zinnikus, I. A Survey on the Use of Synthetic Data for Enhancing Key Aspects of Trustworthy AI in the Energy Domain: Challenges and Opportunities. *Energies* 2024, 17, 1992
- [95] Luo, D., & Pan, W. (2025). Ethical Considerations in Smart Grid Optimization: Promoting Energy Equity and Fairness. *International Theory and Practice in Humanities and Social Sciences*, 2(2), 145–163
- [96] J. Thibodeau, H. Nekoei, A. Ta'ik, J. Rajendran, G. Farnadi, Fairness incentives in response to unfair dynamic pricing (2024). arXiv:2404.14620
- [97] J. Li, M. Motoki, B. Zhang, Balancing fairness and efficiency in energy resource allocations (2024). arXiv:2403.15616
- [98] Y.-G. Lee, J.-Y. Oh, G. Kim, Interpretation of load forecasting using explainable artificial intelligence techniques, *The Transactions of the Korean Institute of Electrical Engineers* 69 (2020) 480–485
- [99] N. Sakkas, S. Yfanti, C. Daskalakis, E. Barbu, M. Domnich, Interpretable forecasting of energy demand in the residential sector, *Energies* 14 (20) (2021)
- [100] S. Pelekis, T. Pountridis, G. Kormpakis, G. Lampropoulos, E. Karakolis, S. Mouzakitis, D. Askounis, Deeptsf: Codeless machine learning operations for time series forecasting, *SoftwareX* 27 (2024) 101758
- [101] M. Shajalal, A. Boden, G. Stevens, Forecastexplainer: Explainable household energy demand forecasting by approximating shapley values using deeplift, *Technological Forecasting and Social Change* 206 (2024) 123588
- [102] S. Sarp, M. Kuzlu, U. Cali, O. Elma, O. Guler, An interpretable solar photovoltaic power generation forecasting approach using an explainable artificial intelligence tool, in: 2021 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT), 2021, pp. 1–5.
- [103] H. Wang, R. Cai, B. Zhou, S. Aziz, B. Qin, N. Voropai, L. Gan, E. Barakhtenko, Solar irradiance forecasting based on direct explainable neural network, *Energy Conversion and Management* 226 (2020) 113487
- [104] M. Chaibi, E. M. Benghoulam, L. Tarik, M. Berrada, A. E. Hmaidi, An interpretable machine learning model for daily global solar radiation prediction, *Energies* 14 (21) (2021).
- [105] X. Cui, M. Lee, C. Koo, T. Hong, Energy consumption prediction and household feature analysis for different residential building types using machine learning and SHAP: Toward energy-efficient buildings, *Energy and Buildings* 309 (2024) 113997
- [106] M. Sun, C. Han, Q. Nie, J. Xu, F. Zhang, Q. Zhao, Understanding building energy efficiency with administrative and emerging urban big data by deep learning in Glasgow, *Energy and Buildings* 273 (2022) 112331
- [107] B. S. Alotaibi, Advancing energy performance efficiency in residential buildings for sustainable design: Integrating machine learning and optimized explainable AI (AIX), *International Journal of Energy Research* 2024 (2024) 6130634.
- [108] P. Rempi, S. Pelekis, A. M. Tzortzis, E. Karakolis, C. Ntanos, D. Askounis, Explainable ai for building energy retrofitting under data scarcity (2025). arXiv:2504.06055.

-
- [109] K. Konhäuser, T. Werner, Uncovering the financial impact of energy-efficient building characteristics with explainable artificial intelligence, *Applied Energy* 374 (2024) 123960
- [110] M. R. Shadi, H. Mirshekali, H. R. Shaker, Explainable artificial intelligence for energy systems maintenance: A review on concepts, current techniques, challenges, and prospects, *Renewable and Sustainable Energy Reviews* 216 (2025) 115668
- [111] S. Sairam, S. Srinivasan, G. Marafioti, B. Subathra, G. Mathisen, K. Bekiroglu, Explainable incipient fault detection systems for photovoltaic panels (2020). arXiv:2011.09843
- [112] J. Dou, X. Wang, Z. Liu, Z. Jiao, Y. Han, Q. Sun, J. He, Multi-objective explainable smart dispatch for integrated energy system based on an explainable mo-rl method, *Comput. Electr. Eng.* 118 (PB) (Sep. 2024).
- [113] H. N. Noura, Z. Allal, O. Salman, K. Chahine, Explainable artificial intelligence of tree-based algorithms for fault detection and diagnosis in grid-connected photovoltaic systems, *Engineering Applications of Artificial Intelligence* 139 (2025) 109503
- [114] "Toward a Comprehensive Framework for Ensuring Security and Privacy in Artificial Intelligence" - William Villegas, Joselin García-Ortiz, 2023
- [115] "State data security backed by Artificial Intelligence and Zero Knowledge Proofs in the context of sanctions and economic pressure" - Abdul Qawi Jalili, Anton Dzyatkovskii, 2023
- [116] The Global Governance of Artificial Intelligence: Next Steps for Empirical and Normative Research, *International Studies Review*, Tallberg et al, 2023
- [117] A systematic review of regulatory strategies and transparency mandates in AI regulation in Europe, the United States, and Canada, Sloane, Mona and Wüllhorst, Elena, 2025
- [118] Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation, Natalia Díaz-Rodríguez et al., *Information Fusion* (2023)
- [119] Human-Centric Versus State-Driven: A Comparative Analysis of the European Union's and China's Artificial Intelligence Governance Using Risk Management, *International Journal of Intelligent Information Technologies*, Arora et al, (2021)
- [120] David Leslie, Christopher Burr, Mhairi Aitken, Josh Cowls, Mike Katell, & Morgan Briggs; With a foreword by Lord Tim Clement-Jones, *Artificial intelligence, human rights, democracy, and the rule of law - A primer* (2021),
- [121] Jørgensen, B.N.; Ma, Z.G. Regulating AI in the Energy Sector: A Scoping Review of EU Laws, Challenges, and Global Perspectives. *Energies* 2025, 18, 2359
- [122] Hannah Ruschemeier, AI as a challenge for legal regulation – the scope of application of the artificial intelligence act proposal, *ERA Forum* (2023)
- [123] Jørgensen, B.N.; Ma, Z.G. Regulating AI in the Energy Sector: A Scoping Review of EU Laws, Challenges, and Global Perspectives. *Energies* 2025, 18, 2359
-

- [124] Jørgensen, B.N.; Ma, Z.G. Regulating AI in the Energy Sector: A Scoping Review of EU Laws, Challenges, and Global Perspectives. *Energies* 2025, 18, 2359
- [125] Hans Graux, Pieter Gryffroy, Magdalena Gad- Nowak, Liesa Boghaert, Data Europe EU, The role of artificial intelligence in the processing and generating new data
- [126] Gangale, F., Mengolini, A.M., Covrig, L., Chondrogiannis, S. and Shortall, R., Making energy regulation fit for purpose. State of play of regulatory experimentation in the EU, EUR 31438 EN, Publications Office of the European Union, Luxembourg, 2023, ISBN 978-92-68-00415-9
- [127] Ranchordas, Sofia and Vinci, Valeria, Regulatory Sandboxes and Innovation-friendly Regulation : Between Collaboration and Capture (January 16, 2024). Sofia Ranchordas & Valeria Vinci, Regulatory Sandboxes and Innovation-friendly Regulation: Between Collaboration and Capture, Italian Journal of Public Law, Vol. 1/2024, Forthcoming , Tilburg Law School Research Paper
- [128] Hannah Ruschemeier, Thinking Outside the Box? Regulatory Sandboxes as a Tool for AI Regulation 2024-04-07 Pre-print, under review
- [129] Arto Lanamäki, Karin Väyrynen, Fanny Vainionpää, Heidi Hietala, Erkki Tervo, Alex Moltzau, Sophie Weerts, Demian Niemeyer, What to Expect from the Upcoming EU AI Act Sandboxes: Panel Report, Digital Society (2025)
- [130] Leslie, D. (2019). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute
- [131] Leslie, D., Burr, C., Aitken, M., Cows, J., Katell, M., and Briggs, M. (2021). Artificial intelligence, human rights, democracy, and the rule of law: a primer. The Council of Europe
- [132] Understanding Relations Between Perception of Fairness and Trust in Algorithmic Decision Making, Jianlong Zhou et al., The International Conference on Behavioral and Social Computing (BESC 2021)
- [133] Smuha, Nathalie A., The Work of the High-Level Expert Group on AI as the Precursor of the AI Act (October 01, 2024). Available at SSRN

Annex II: Requirements gathering - questionnaires

The requirements' gathering phase has been performed by means of questionnaires and interviews. In this section, the different questionnaires for the different types of stakeholders are reported. The filled questionnaires are available as annexes to this deliverable.

AI Developers

AI Developers especially SMEs and startups, software engineers of solutions for AI, HPC, big data in Energy sector domains. Link to online questionnaire [here](#).

Section 1: General Information

1. Name of your organization:
2. Country:
3. Contact person (name, email, role):
4. Type of organization:
 - ☐ Startup
 - ☐ Small or Medium Enterprise (SME)
 - ☐ Large Enterprise
 - ☐ Research Institution/University
 - ☐ Public Sector Organization (NGO)
 - ☐ Other (please specify):
5. Industry sector focus (select all that apply):
 - ☐ Energy
 - ☐ Energy Supplier
 - ☐ Renewable Energy Provider
 - ☐ Nuclear energy
 - ☐ Oil & Gas
 - ☐ Electricity Transmission & Distribution
 - ☐ Energy Storage & Battery Technologies
 - ☐ Energy Efficiency & Demand Management (e.g. smart grids, energy monitoring, and efficiency solutions)
 - ☐ Hydrogen & Alternative Fuels
 - ☐ Carbon Capture & Emissions Reduction
 - ☐ Energy Consulting & Services
 - ☐ ICT
 - ☐ AI and machine learning
 - ☐ Cloud computing and data centres
 - ☐ Cybersecurity
 - ☐ Telecommunications
 - ☐ Software and IT services
 - ☐ Other (please specify):

Section 2: AI Solution Overview

6. Do you develop of plan to develop an AI application for the energy sector?

☐ Yes

☐ No

If yes, please describe it:

7. What is the main objective of your AI solution?

☐ Predictive maintenance (e.g. detect equipment failures before they happen)

☐ Energy Optimization, to enhance efficiency in power generation, distribution or consumption

☐ Grid Management & Load Forecasting

☐ Cybersecurity & Threat Detection

☐ Renewable Energy Forecasting (e.g. predicting solar and wind power generation)

☐ Smart Energy Trading & Market Optimization

☐ Carbon Emissions Monitoring & Reduction

☐ Other (please specify):

8. What AI technologies/methods are you using?

☐ Machine learning

☐ Digital twins

☐ Large language models

☐ Computer vision

☐ Reinforcement Learning

☐ AI Agents

☐ Other (please specify):

9. What stage of development is your AI solution currently in?

☐ Conceptualization

☐ Prototype

☐ Pilot Testing

☐ Market-ready

☐ Yes

☐ No

If yes, where and how?

10. Have you previously tested your AI solution in a real-world environment?

Section 3: TEF Requirements & Expectations

11. What are your main objectives for using the EnergyGuard TEF? -

☐ Access to real-world energy data

☐ Testing and validation under real conditions

☐ AI performance optimization

☐ Cybersecurity and trustworthiness assessment

☐ Compliance with EU AI regulations

☐ Use High Performance Computing (HPC) resources

☐ Other (please specify):

12. What type of data would you need for your testing? -

☐ Smart grid data

☐ Renewable energy production data

☐ Building energy consumption data

- ☐ Weather & climate data
 - ☐ Equipment data (e.g. operational status, vibrations, temperature)
 - ☐ Energy Market & pricing data
 - ☐ Carbon Emissions & Sustainability Data
 - ☐ Battery & Energy Storage Data
 - ☐ Industrial Process Energy Usage
 - ☐ Cybersecurity & Network Data
 - ☐ Other (please specify):
13. Are there specific AI regulatory or compliance requirements that you need to meet?
- ☐ Yes (please specify)
 - ☐ No
14. What are your expectations regarding EnergyGuard TEF's web user interface (e.g. viewing the list of available facilities, viewing the list of available data, etc.)?
15. What are your expectations regarding EnergyGuard TEF's APIs (e.g. performance requirements, size of data sets, etc.)?
16. What are your expectations regarding EnergyGuard TEF's support services?
- ☐ Technical guidance and consulting
 - ☐ Legal and compliance support
 - ☐ HPC infrastructure access -
 - ☐ Networking and collaboration opportunities
 - ☐ Other (please specify):

Section 4: Additional Information

17. What are the biggest challenges you face in AI testing and validation?
- ☐ Lack of High-Quality & Representative Data
 - ☐ Understanding and justifying AI decisions
 - ☐ Scalability & Performance Testing
 - ☐ Robustness & Security Risks
 - ☐ Compliance with Regulatory Requirements (e.g., GDPR, EU AI Act)
 - ☐ Integration with Real-World Systems (e.g., legacy infrastructure, APIs)
 - ☐ Lack of Standardized Evaluation Frameworks or Metrics
 - ☐ Limited Access to Testing Infrastructure (e.g., compute resources, sandboxes)
 - ☐ Other (please specify):
18. Do you have any specific requirements for AI model interoperability?
- ☐ Standardized Data Formats & APIs
 - ☐ Cross-Platform Compatibility (hardware, software, cloud)
 - ☐ Model Portability & Reusability
 - ☐ Compliance with Industry Standards
 - ☐ Other (please specify):
19. Are you interested in collaborating with other AI developers, researchers, or energy sector stakeholders?
- ☐ Yes
 - ☐ No
20. Any other comments or additional requirements?

Energy Production and Energy Systems (EPES) value chain stakeholders

Generators, suppliers, TSOs, DSOs, hydrogen players, aggregators, building operators, local energy communities, suppliers, ESCOs, facility manager, RES plants O&M. Link to online questionnaire [here](#).

Section 1: General Information

1. Name of your organization:
2. Country:
3. Contact person (name, email, role):
4. What is your organization's primary role in the EPES sector? (Select one or more options)
 - ☐ Energy generator
 - ☐ Energy supplier
 - ☐ Transmission System Operator (TSO)
 - ☐ Distribution System Operator (DSO)
 - ☐ Hydrogen sector player
 - ☐ Aggregator
 - ☐ Building operator
 - ☐ Local energy community
 - ☐ Energy Service Company (ESCO)
 - ☐ Facility manager
 - ☐ RES (Renewable Energy Sources) plants O&M operator
 - ☐ Other (please specify):
5. What energy technologies is your organization primarily responsible for? (Select one or more options)
 - ☐ Renewable energy sources (RES)
 - ☐ Wind
 - ☐ Solar PV
 - ☐ Hydro
 - ☐ Biomass
 - ☐ Other RES (please specify):
 - ☐ Non-renewable energy sources (non-RES)
 - ☐ Gas
 - ☐ Coal
 - ☐ Nuclear
 - ☐ Waste-to-energy
 - ☐ Diesel / Oil
 - ☐ Other non-RES (please specify):
 - ☐ Energy storage systems
 - ☐ Hydrogen and related technologies
 - ☐ Demand management and aggregation systems
 - ☐ Electricity grid technologies
 - ☐ Generation (please indicate Type A/B/C/D if applicable):
 - ☐ Transmission

- ☐ Distribution
- ☐ Heat grid technologies
- ☐ Gas grid technologies
- ☐ Other (please specify):

Section 2: Challenges and Needs

6. What are the new solutions and innovative technologies that you are willing to adopt or are interested in, and what are the main challenges your organization faces in adopting them?
 - ☐ AI for forecasting and optimization
 - ☐ AI for predictive maintenance and asset management
 - ☐ AI for demand-side management and customer interaction
 - ☐ AI for operational security and anomaly detection
 - ☐ AI-based digital twins and simulation tools
 - ☐ AI for Distributed Energy Resources and storage integration
 - ☐ Other AI-related technologies (please specify)
 - ☐ Other innovative solutions (non-AI) (please specify):
7. What are the key barriers to adopting these innovative Technologies in the EPES sector? (Select up to 3 options)
 - ☐ High costs
 - ☐ Complex regulations
 - ☐ Interoperability with existing systems
 - ☐ End-user acceptance
 - ☐ Lack of financial incentives
 - ☐ Other (please specify):
8. What are your organization's main needs to improve operational efficiency and sustainability?

Section 3: Innovation and Collaboration

9. Does your organization actively participate in innovation or research projects?
 - ☐ No
 - ☐ Yes, please specify most recent ones:
10. What aspects do you consider a priority in developing new solutions and innovative technologies? (Select up to 3 options)
 - ☐ Greater integration of renewables
 - ☐ Reduction of grid losses
 - ☐ Digitalization and smart grids
 - ☐ Increased flexibility in demand management
 - ☐ More efficient storage systems
 - ☐ Other (please specify):
11. How important is collaboration with other EPES value chain actors for your organization?
 - ☐ Very important
 - ☐ Important
 - ☐ Not very important
 - ☐ Not important at all

12. What digital tools or platforms does your organization use to optimize energy/asset/system management?
13. Does your organization currently use artificial intelligence (AI) solutions in energy/asset/system management, and in which areas?
 - ☐ Forecasting energy demand and supply
 - ☐ Predictive maintenance of energy infrastructure
 - ☐ Optimization of grid operations and flexibility
 - ☐ Energy trading and market analysis
 - ☐ Automation of customer energy services
 - ☐ Other (please specify):
 - ☐ We do not use AI solutions
14. What are the main challenges your organization faces in adopting AI-based solutions for energy systems? (Select up to 3 options)
 - ☐ Lack of internal expertise
 - ☐ High implementation costs
 - ☐ Data availability and quality issues
 - ☐ Regulatory and compliance constraints
 - ☐ Integration with existing infrastructure
 - ☐ Other (please specify):

Section 4: Regulation and Sustainability

15. What regulations do you consider most impactful for your organization's activities?
16. Does your organization have specific decarbonization or sustainability goals?
 - ☐ No
 - ☐ Yes, please specify which ones:
17. What incentives or regulatory support would you find useful to facilitate the energy transition?

Section 5: Final Considerations

18. What suggestions would you give to improve collaboration between different EPES value chain actors?
19. Any other comments or additional requirements?

Lawyers and Legal system entities

Lawyers, Law firms, judges, magistrates, Law Enforcement Agencies

1. Name of your organization:
2. Country:
3. Contact person (name, email, role):
4. Please indicate the type of organization you represent:
 - ☐ Lawyer
 - ☐ Law firm or legal consultancy
 - ☐ Judge
 - ☐ Magistrate
 - ☐ Law enforcement agency

- ☐ Academic/legal research institution
- ☐ Other (please specify):
5. Are you currently involved in legal or regulatory work related to AI or energy systems?
- ☐ Yes
- ☐ No
- ☐ Not directly, but with interest
6. From your legal perspective, what are the most critical risks in deploying AI in the energy sector?
(Select up to 3)
- ☐ Lack of accountability and liability frameworks
- ☐ Data protection and privacy issues
- ☐ Discrimination and algorithmic bias
- ☐ Lack of transparency in AI decision-making
- ☐ Insufficient regulatory guidance
- ☐ Other (please specify)
7. Are current EU or national legal frameworks adequate to regulate AI in energy use cases?
- ☐ Yes
- ☐ Partially
- ☐ No
8. In your view, what type of legal instruments or mechanisms are needed to ensure trustworthy AI in the energy sector?
9. How relevant are the following legal areas in the context of AI-based energy systems? (Rate from 1 = Not relevant to 5 = Very relevant)
- ☐ Data protection (e.g. GDPR)
- ☐ Consumer rights and protection
- ☐ Competition and market access
- ☐ Liability and accountability
- ☐ Cybersecurity law
- ☐ Environmental law
10. Do you believe a sector-specific regulatory sandbox (e.g. for AI in energy) would be beneficial?
Why or why not?
11. Would your organization be interested in contributing legal expertise to the EnergyGuard project
(e.g. through workshops, white papers, or advisory roles)?
- ☐ Yes
- ☐ No
- ☐ Maybe (please specify conditions)

Industry Associations & Technology Clusters

European initiatives and clusters (European AI Alliance, AIOD, BDVA, Adra-e, IDSA, AIOTI), research communities, associations, federations (IMS, EFFRA, IFIP, IEEE, NEM).

1. Name of your organization:
2. Country:
3. Contact person (name, email, role):
4. Please indicate the type of organization you represent:
 - ☐ Technology/Innovation Cluster
 - ☐ Industry Association
 - ☐ Research Community or Federation
 - ☐ Standardization Body
 - ☐ Other (please specify):
5. Which sectors or domains does your organization primarily focus on? (Multiple selection allowed)
 - ☐ Artificial Intelligence
 - ☐ Energy and Utilities
 - ☐ Digital Infrastructure and Data Spaces
 - ☐ Cybersecurity
 - ☐ Manufacturing and Industry 4.0
 - ☐ Other (please specify)
6. How does your organization support the adoption of AI technologies in the energy sector?
7. From your network's perspective, what are the main barriers to adopting AI in energy systems? (Select up to 3)
 - ☐ Lack of trust in AI solutions
 - ☐ Data availability and interoperability
 - ☐ Fragmentation of standards and platforms
 - ☐ Regulatory and ethical concerns
 - ☐ Skills and workforce readiness
 - ☐ Other (please specify)
8. Which of the following areas should be prioritized in future AI testing and experimentation for the energy sector? (Select up to 3)
 - ☐ Grid and infrastructure optimization
 - ☐ Energy communities and prosumer engagement
 - ☐ Hydrogen systems and RES integration
 - ☐ Cybersecurity and resilience testing
 - ☐ AI transparency, explainability, and ethics
 - ☐ Other (please specify)
9. Is your organization involved in any EU-funded initiatives or working groups related to AI, energy, or data ecosystems?
 - ☐ No
 - ☐ Yes, please specify which ones:
10. How could your organization contribute to the EnergyGuard ecosystem (e.g., dissemination, stakeholder engagement, policy guidance, access to networks)?

11. Would your organization be interested in participating in the project's future workshops or other initiatives?

☐ No

☐ Yes

☐ Maybe, please specify conditions:

Investors

Banks, RES/EE Investors/Venture capitalists, Crowdsourcing platforms.

1. Name of your organization:

2. Country:

3. Contact person (name, email, role):

4. What type of investment entity do you represent?

☐ Commercial Bank

☐ Venture Capital Firm

☐ Private Equity or Investment Fund

☐ Crowdfunding or Crowdfunding Platform

☐ Public Funding Agency

☐ Other (please specify):

5. How familiar are you with investments in the energy or cleantech sector?

☐ Not familiar

☐ Somewhat familiar

☐ Actively involved in related investments

6. What is your current level of interest in funding projects that involve the use of AI in the energy sector?

☐ High

☐ Moderate

☐ Low

☐ Not interested now

7. Which of the following factors are most important when evaluating such investments? (Select up to 3)

☐ Market size and growth potential

☐ Technological maturity and scalability

☐ Regulatory environment and risks

☐ Time to market / Return on Investment (ROI)

☐ ESG impact (Environmental, Social, Governance)

☐ Quality of the management/technical team

☐ Other (please specify)

8. What are your main concerns when evaluating investments in AI-based energy solutions?

9. Would sector-specific AI testing facilities, like those offered by EnergyGuard, increase your confidence in the viability of emerging technologies?

☐ Yes

☐ No

☐ Maybe, please specify:

10. What kind of business models do you consider most promising for AI in energy? (Select one or more)

☐ SaaS (Software as a Service) platforms

☐ Hardware + AI bundled products

☐ Platform-based marketplaces

☐ Subscription models

☐ Data-as-a-service

☐ Other (please specify)

11. Would you be interested in learning more about early-stage companies or use cases emerging from EnergyGuard?

☐ Yes, for potential investment

☐ Yes, for partnership or strategic scouting

☐ No

☐ Maybe, please specify:

Researchers and Academia

Individuals engaged in research initiatives and/ or working in research/academic institutes especially in ML and AI (including AIOD).

1. Name of your organization:

2. Country:

3. Contact person (name, email, role):

4. Please indicate the type of organization you represent:

☐ University

☐ Public Research Institute

☐ Private Research Institute

☐ European Research or Innovation Network

☐ Professional and Scientific Association

☐ Digital Innovation Hubs and Collaborative Labs

☐ Other (please specify):

5. What is your main area of research? (Select one or more)

☐ Machine Learning / Deep Learning

☐ Explainable / Trustworthy AI

☐ Smart Energy Systems

☐ Digital Twins

☐ Data Science / Big Data

☐ Cybersecurity

☐ Other (please specify):

6. How familiar are you with applying AI/ML in the energy domain?

☐ Not familiar

☐ Basic knowledge

☐ Moderate experience

- ☐ Extensive experience (e.g. leading or contributing to projects)
7. Are you currently involved in any research projects related to AI in energy systems?
- ☐ No
- ☐ Yes, please briefly describe your role and the scope:
8. Which topics do you believe require more research and experimentation to enable trustworthy AI adoption in energy systems?
9. What types of research support or infrastructure would be most useful for your work? (Select up to 3)
- ☐ Access to real-world energy datasets
- ☐ Large-scale AI testing environments
- ☐ Simulation tools or digital twins
- ☐ Evaluation metrics for trustworthy AI
- ☐ Collaboration platforms with industry
- ☐ Other (please specify):
10. What challenges do you face in transferring your research to real-world energy applications? (Open-ended)
11. Would you be interested in contributing to EnergyGuard through scientific collaboration (e.g. piloting, testing, co-authoring papers, etc.)?
- ☐ No
- ☐ Yes
- ☐ Maybe, please specify conditions:
12. Do you have any recommendations for ensuring alignment between the EnergyGuard infrastructure and academic research needs?

Policy Makers & Standardisation Organisations (SDO)

Policymakers at any level like EC Directorates and Units, Ministries and Governments, Regulatory Agencies, Standardisation Organisations (CEN, ISO, ETSI, etc.).

1. Name of your organization:
2. Country:
3. Contact person (name, email, role):
4. Please indicate the type of organization you represent:
 - ☐ EU Directorate/Agency
 - ☐ National/Regional Ministry or Government
 - ☐ Regulatory Agency (Energy, Data, Digital, etc.)
 - ☐ Standardisation Organisation (e.g. CEN, ISO, ETSI)
 - ☐ Other (please specify)
5. How relevant is the intersection of artificial intelligence and energy in your current policy or regulatory agenda?
 - ☐ Very relevant
 - ☐ Moderately relevant
 - ☐ Low priority
 - ☐ Not currently addressed
6. Which areas of policy, regulation or standardisation are most impacted using AI in energy systems? (Select up to 3)
 - ☐ Data protection and cybersecurity
 - ☐ Energy market operations and grid management
 - ☐ Environmental and climate policy
 - ☐ Consumer protection and rights
 - ☐ Certification and standardisation of AI tools
 - ☐ Other (please specify)
7. What are the main regulatory or standardisation gaps you perceive regarding AI deployment in energy applications?
8. Do you see value in having sector-specific AI testing and experimentation facilities, like EnergyGuard, to support policy validation or regulatory sandboxes?
 - ☐ Yes
 - ☐ No
 - ☐ Not surePlease explain your answer (optional)
9. How important is alignment between European AI regulations (e.g. the AI Act) and energy sector-specific standards or certifications?
 - ☐ Very important

- ☐ Important
- ☐ Neutral
- ☐ Not important
- ☐ Not sure

10. Please specify 3-5 key points requiring alignment between regulation and standardization.
11. Would your institution be interested in contributing to policy shaping or standardisation activities related to the EnergyGuard initiative?
 - ☐ Yes
 - ☐ No
 - ☐ Maybe, please specify conditions:
12. Which tools or mechanisms would you recommend ensuring regulatory compliance and trustworthy AI adoption across the energy value chain?

General Public

General public, Civil society representatives, anyone interested in EnergyGuard.

1. Name of your organization:
2. Country:
3. Contact person (name, email, role):
4. Please indicate the type of organization you represent:
 - ☐ Private citizen / energy consumer
 - ☐ Local Government
 - ☐ Small or Medium Enterprise
 - ☐ Large Enterprise
 - ☐ Other (please specify):
5. Have you heard about artificial intelligence (AI) being used in the energy sector?
 - ☐ No
 - ☐ Yes
 - ☐ Not sure
6. In your opinion, what are the most important benefits that AI could bring to the energy sector? (Select up to 3)
 - ☐ Lower energy bills
 - ☐ More efficient energy systems
 - ☐ Better integration of renewable energy
 - ☐ Faster response to outages or issues
 - ☐ Easier energy management at home
 - ☐ Don't know / No opinion
7. What concerns (if any) do you have about the use of AI in energy systems?
8. How important is it for you to know that AI systems in energy are tested for safety, transparency, and fairness before being used?
 - ☐ Very important

- ☐ Important
☐ Neutral
☐ Not important
☐ Don't know / No opinion
9. Would you be interested in receiving updates about new tools or services developed through EnergyGuard?
- ☐ Yes
☐ No
☐ Maybe, please specify conditions:
10. Which formats would you prefer to receive information in? (Select one or more)
- ☐ Website updates
☐ Social media posts
☐ Email newsletters
☐ Public events / community talks
☐ Videos / short explainers
11. In your view, who should benefit the most from AI-based energy innovations?
12. Would you be willing to take part in future public consultations, surveys or events related to this project?
- ☐ Yes
☐ No
☐ Maybe, please specify conditions:

Pilot use cases

Pilot use cases are the experiments to run or the apps to develop with the use of the TEF facilities.

Overview	
Owner of the Pilot	(name of partner)
Use Case Acronym	(Unique acronym to identify the use case) – PX_UCY
Document Version	(Version of the information about the data source)
Scope / Objectives of Use Case-Experiment	
Current Situation	
As-is Situation - Describe the current situation for the development of your experiment	
Use Case	
Diagram of Use Case	
(picture with a diagram illustrating the use case)	

Primary Actor	Who is the main actor that this use case represents
Secondary Actors	Who is/are the secondary actors that this use case represents (if applicable)
AI techniques / Models	Linked with above tables Usage of AI techniques: ML/DL Models: Regression Analysis, Decision trees, ANN, ... Data sets (Nature, size, format, Data Ingestion, Curation, Preprocessing, ...), (Distributed/Federated) supervised/unsupervised training, inference constraints Hw/Sw Acceleration, etc.
Equipment / Hardware used	Linked with above tables Hardware / equipment used in the workflows examples Computing technologies: CPU, GPU, FPGA, DSP, etc. Interconnection Fabric: IB; Ethernet, etc. Storage devices (type + capacity) Data volume to be manipulated in memory and/or storage
Software used	Linked with above tables Software used in the workflows: Maturity: Prototype, industrial exploitation, commercialization, etc. Sources: proprietary development, open source, etc. Parallelization and Programming Models (OpenMP, MPI, OpenCL, CUDA, etc.) Open-Source Libraries, Independant Software Vendor products, ... OS / programming language / Compilers
Use Case Description	
To-be Situation – Elaborate on the future envisaged business process/use case	
Technical Challenges	
User Perspective	
Perspective of the Users (what is the advantage of the TEF for the users, when compared with the current situation?)	
System Perspective	
Perspective of the system (how do you envisage that the existing platforms interact with the TEF platform?)	
Details	

Business Goals/Benefits	
Why is the use case executed? What are the benefits?	
Application	What system or application does this pertain to
Data Preconditions	What data preconditions must be met before this use case can start
System Preconditions	What system preconditions must be met before this use case can start
User Preconditions	What user preconditions must be met before this use case can start
Trigger	What event triggers this use case
Challenges/Limitations	What can constitute an obstruction to the use case
Expected results - KPIs	
<i>Actual performance (e.g. execution time), limitations, bottlenecks, etc.</i> Perspectives of evolution: <i>Computing power, computing technology (CPU, GPU, FPGA), new algorithms, Application of AI, etc.</i> Quantification of needs: <i>Speedup, increase of data volume, improvement of accuracy, etc.</i>	
User Journey, stories, requirements, data sources, models, assets to be shared	
Link to users, user journeys, use cases, etc. above	

Table 9: Pilot use cases questionnaire