Horizon Europe Framework Programme (HORIZON)

HORIZON-CL4-2023-HUMAN-01-01: Efficient trustworthy AI - making the best of data (Adra)

Grant Agreement: 101135826

AI-Ops Framework for Automated, Intelligent and Reliable Data/AI Pipelines Lifecycle with Humans-in- the-Loop and Coupling of Hybrid Science-Guided and AI Models

WP1: Automated AI Pipeline Lifecycle Management Framework

D1.2: AI-DAPT Automated AI Pipeline Design and Technical Requirements

Deliverable Leader: UCY

Due Date: M12 (December 31st, 2024)

Dissemination Level: Public

Version: D1.2

Short Abstract

D1.2: AI-DAPT Automated AI Pipeline Design and Technical Requirements presents the foundational work completed in AI-DAPT's WP1, including: (i) the updated Research Agenda, building on the Scientific and Technological Radar from D1.1 and addressing key AI topics, (ii) updated descriptions of the Demonstrators, their data sources, and relevant open datasets, (iii) the initial design of the envisioned data and AI pipeline for AI-DAPT, (iv) end-to-end usage scenarios aligned with user needs identified in D1.1, (v) the technical requirements, both functional and non-functional, that guide the platform's development, (vi) the Minimum Viable Product (MVP) framework for the platform, and (vii) the ethical and legal requirements tailored to each Demonstrator.

Further Information: www.aidapt.eu

Disclaimer. The views represented in this document only reflect the views of the authors and not the views of the European Union. The European Union is not liable for any use that may be made of the information contained in this document. Furthermore, the information is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose. The user of the information uses it at its sole risk and liability.

Partners

	ATHINA- Erevnitiko Kentro Kainotomias stis Technologies tis Pliroforias, ton Epikoinonion kai tis Gnosis (ATHENA)	Greece
	Fraunhofer Gesellschaft zur Forderung der Angewandten Forschung ev (FRAUNHOFER)	Germany
	South East Technological University (SETU)	Ireland
	Uninova-Instituto de Desenvolvimento de Novas Tecnologias (UNINOVA)	Portugal
	Universitat Politècnica de Catalunya (UPC)	Spain
	Suite5 Data Intelligence Solutions Limited (Suite5)	Cyprus
	MCS Datalabs (MCS)	Germany
	Eyfyia gia Epicheiriseis Etaireia Periorismenis Evthinis - Intelligence for Business Ltd (WITSITE)	Greece
	University of Cyprus (UCY)	Cyprus
	S&D Consulting Europe SRL (S&D)	Italy
	Gioumpitek Meleti Schediasmos Ylopoiisi kai Polisi Ergon Pliroforikis Etaireia Periorismenis Efthynis (UBITECH)	Greece

CHARITÉ UNIVERSITÄTSMEDIZIN BERLIN	Charite - Universitaetsmedizin Berlin (CHARITE)	Germany
BIBA	BIBA - Bremer Institut fuer Produktion und Logistik Gmbh (BIBA)	Germany
OHS engineering	OHS Engineering Gmbh (OHS)	Germany
Zéniθ	Etaireia Promitheias Aeriou Thessalonikis - Thessalias Monoprosopi Anonymos Etaireia (ZENITH)	Greece
domx IOT TECHNOLOGIES	DOMX Idiotiki Kefalaiouchiki Etaireia (DOMX)	Greece
MADE Competence Center i4.0	Made Scarl (MADE)	Italy
INTELLIMECH® CONSORZIO PER LA MECCATRONICA	Consorzio Intellimech (IMECH)	Italy

Document Log

Contributors	ATHENA (Dimitris Kyriakopoulos, Dimitris Skoutas, Eleni Lavasa, Giorgos Giannopoulos, Theodore Dalamagas, Vasilis Gkolemis) BIBA (Robert Hellbach, Karl Hribernik) CHARITE (Bettina Schuppelius, Marta Csanalosi Artigas) DOMX (Stratos Keranidis, Viktor Daropoulos) MCS (André Tabone, Fihmi Mousa) OHS (Carl Hans) SUITE5 (Nefeli Bountouni, Sotiris Koussouris) S&D (Marina Cugurra) UCY (Charalambos Lambri, George Ioannou, George Pallis, Joanna Georgiou, Marios Dikaiakos) UNINOVA (André Grilo, Paulo Figueiras) UPC (Adrian Asensio, Gladys Utrera, Xavi Masip) ZENITH (Dimitris Bimpikas)
Internal Reviewer 1	George Pallis - UCY
Internal Reviewer 2	Stratos Keranidis - DOMX
Type	Report
Delivery Date	M12

History

Versions	Description
v0.1	Table of Contents
v0.2	First draft of the document
v1.0	Second draft of the document
v2.0	Final Version

Executive Summary

This document, Deliverable D1.2 - *AI-DAPT Automated AI Pipeline Design and Technical Requirements*, outlines foundational work necessary to bring AI-DAPT from concept to implementation. Prepared as part of AI-DAPT's WP1, this deliverable establishes the groundwork for both ongoing and future efforts. WP1 will continue through M33, with iterative updates planned across all tasks. This deliverable not only charts a clear path for the remainder of WP1 but also sets the stage for subsequent technical work packages, specifically WP2, WP3, and WP4.

Deliverable D1.2 consolidates progress from tasks T1.1, T1.2, T1.4, and T1.5, contributing key outputs that advance AI-DAPT's scientific and technological objectives. These include:

- An updated research agenda that explores critical advancements in AI-related fields, including data processing, synthetic data generation, explainable AI, bias detection, science-guided AI, and adaptive AI. This agenda drives research efforts across WP2, WP3, and WP4, ensuring alignment with the project's goals.
- A comprehensive analysis of the four AI-DAPT Demonstrator scenarios, encompassing updated descriptions of data sources, open datasets, and evolving user needs tailored to each scenario.
- An initial design for AI-DAPT's end-to-end data and AI pipeline, detailing all key phases of the pipeline lifecycle—from data preparation and nurturing to model delivery and optimization. This design establishes a baseline for implementing adaptable and automated workflows within the platform.
- Stakeholder-driven end-to-end usage scenarios, capturing functional journeys and expectations that align pipeline operations with technical requirements and user-centric objectives.
- Detailed functional and non-functional technical requirements that underpin the iterative development of the platform's Minimum Viable Product (MVP).
- An expanded ethics analysis, addressing ethical constraints and legal considerations. This includes tailored frameworks and principles for each Demonstrator, focusing on human-in-the-loop (HITL) processes, regulatory compliance, and data protection.

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 AI-DAPT Project Overview

Today, Artificial Intelligence (AI) has paved a long way since its inception and has started experiencing exponential growth across various industries and shaping our world in ways that were once thought impossible. As AI transitions from research to deployment, leveraging the appropriate data to develop and evaluate AI models has evolved into one of its greatest challenges. Data are in fact the raw material and the most indispensable asset fuelling much of today's progress in AI, generating previously unattainable insights, assisting more evidence-based decision-making, and bringing tangible business/economic benefits and innovation to all involved stakeholders. However, despite their instrumental role in determining performance, fairness, and robustness of AI systems, data is often an under-valued and de-glamorised aspect of AI while a data-centric focus is typically lacking in the current AI research.

AI-DAPT aims to deliver an innovative and impactful research agenda that will provide tangible benefits to a variety of stakeholders that struggle with making AI services. Seeking to reinstate the pure data-related work in its rightful place, and reinforcing the generalizability, reliability, trustworthiness, and fairness of Al solutions, AI-DAPT vision relies on the implementation of an AIOps framework to support and automate AI pipelines that continuously learn and adapt based on their context. It enables proper purposing, collection, documentation, (bias) valuation, annotation, curation and synthetic generation of data, while keeping humans-in-the-loop across five axis: (i) Data Design for AI, (ii) Data Nurturing for AI, (iii) Data Generation for AI, (iv) Model Delivery for AI, (v) Data-Model Optimization for AI.

AI-DAPT brings forward a two-fold data-centric mentality in AI:

- Data: AI-driven automation for data pipelines based on Explainable AI (XAI) techniques as well as synthetic data generation and observability.
- Model: Automation on AI model building and hybrid science-AI solutions, bringing together data-driven AI models and science-based (first-principle) models that build on high-quality data.

Bridging the gap between data-centric and model-centric AI, AI-DAPT will turn over a new page in trustworthy AI and will nurture an ecosystem involving all AI and data value-chain stakeholders. The aim is to enhance their prosperous collaboration to deliver and apply innovative AI-driven methods that rely on smart and dynamic end-to-end automation of data, AI training/inference pipelines in the cloud-edge computing continuum.

To demonstrate the actual innovation and added value that can be derived through the AI-DAPT scientific advancements, the AI-DAPT results will be validated in two ways:

- By applying them to tackle real-world challenges in four key industries: (1) Health, (2) Robotics, (3) Energy, and (4) Manufacturing.
- By integrating them into various AI solutions, whether open source or commercial, already present in the market.

1.2 Deliverable Purpose and Scope

This deliverable, D1.2, documents the work conducted under WP1, specifically within tasks T1.1, T1.2, T1.4, and T1.5. Overall, it provides the following:

- An updated research agenda that advances AI-DAPT's key scientific and technological goals, including an in-depth exploration of areas such as data processing, synthetic data generation, explainable AI, bias detection, and science-guided AI, supporting the future and ongoing technical developments and research across WP2, WP3, and WP4 (from T1.4).

- A comprehensive overview of the four Demonstrator use-cases, including an updated version of the detailed analyses of relevant data sources and needs specific to each scenario (from T1.2).
- An initial end-to-end pipeline design for data and AI processes, outlining each pipeline phase necessary for AI-DAPT implementation and establishing baseline workflows for platform operations (from T1.4).
- The initial end-to-end usage scenarios that reflect key stakeholders' roles and the functional journey for AI-DAPT across different application contexts (from T1.2).
- Technical requirements, both functional and non-functional, that will serve as the basis for the platform's Minimum Viable Product (MVP) and subsequent stages of development (from T1.5).
- An expanded ethics analysis, detailing the ethical and legal requirements for each Demonstrator, considering updated feedback on ethics and data protection from Demonstrator participants and following the ethics-by-design principles (from T1.1).

This deliverable represents a significant step forward in fulfilling the objectives of WP1. It sets a strong foundation for the rest of the project, with iterative improvements expected in all related tasks until M33.

1.3 Impact and Target Audiences

D1.2 delivers a foundational analysis that supports not only the tasks within WP1 but also extends its influence across WP2, WP3, and WP4. This deliverable defines critical concepts of the AI-DAPT platform, including (i) the envisioned AI and data pipeline, (ii) usage scenarios, (iii) technical requirements, and (iv) the Minimum Viable Product (MVP). Together, these elements will shape the design of the AI-DAPT architecture and provide guidance for its implementation.

Furthermore, D1.2 plays a pivotal role in WP5, which focuses on the Demonstrators. It offers detailed descriptions of each Demonstrator, an updated analysis of the state of the art tailored to each scenario, and a refined overview of the relevant data sources and needs.

The target audience includes all technical partners engaged in the AI-DAPT project, the Demonstrators, and external stakeholders interested in the development and application of AI pipelines and technologies.

1.4 Deliverable Methodology

The creation of Deliverable D1.2 heavily depended on input from the Demonstrators, obtained through questionnaires, dedicated workshops, and interviews. These collaborative efforts played a crucial role in shaping the deliverable by providing a deeper understanding of the Demonstrators' current operations and their evolving needs.

One of the key outcomes of the latter engagements was the provision of updated descriptions of the Demonstrators' data sources. This was important because it offered a clearer picture of the types of data they work with, the challenges they face in data management, and how they use this data in their current processes. These insights were instrumental in identifying potential gaps or inefficiencies in their existing AI systems and laying the groundwork for improvements.

Furthermore, through these discussions, a comprehensive view of the Demonstrators' current AI pipeline processes was identified. This helped to document their existing workflows, including the tools, techniques, and methodologies they are currently using. Understanding the state of their AI pipelines was essential for developing a vision of how these processes could be enhanced and integrated into the broader AI-DAPT framework, which aims to provide a more advanced, adaptive, and efficient AI ecosystem.

In addition to these technical insights, the project team introduced new questionnaires designed to capture the ethical and legal considerations that the Demonstrators face in their work. This was a critical step, as it ensured that the development of AI solutions within the AI-DAPT framework aligns with the project's core principles and objectives, particularly in terms of ensuring ethical AI practices and complying with legal requirements. These new questionnaires helped to clarify the specific concerns of the Demonstrators regarding data privacy, security, transparency, and fairness, all of which are essential components for building trust in AI systems. Based on these, a detailed analysis of the legal requirements was also drafted, and will guide all subsequent actions in this project.

1.5 Dependencies in AI-DAPT and Supporting Documents

D1.2 is based on questionnaires completed by partners involved in the Demonstrator cases, regarding ethical and legal considerations. It also incorporates insights gathered from various workshops and interviews held over the past six months.

This document will serve as a foundation for future deliverables, not only within WP1 (D1.3) but also across all technical and demonstration work packages, including WP2, WP3, WP4, and WP5.

1.6 Document Structure

The rest of the document is organised as follows:

Section 2 elaborates on the updated version of the AI-DAPT research agenda, highlighting advancements in AI-focused areas including data processing, synthetic data generation, explainable AI, bias detection, and science-guided AI. This research agenda will drive iterative progress across all technical work packages, whilst demonstrating the research topics that AI-DAPT will focus on.

Section 3 provides a comprehensive overview of each pilot, including the latest updates on the data sources they intend to leverage within the AI-DAPT platform. Additionally, it identifies relevant open datasets for each pilot, which can serve as baseline resources to support the platform’s development.

In **Section 4**, introduces a preliminary end-to-end data & AI pipeline design for AI-DAPT based on the pilots' envisioned process. This includes a step-by-step breakdown of phases essential to the pipeline lifecycle, from data design and nurturing to model delivery and optimization, laying the groundwork for seamless integration and adaptability across various AI-DAPT applications.

Section 5 outlines the usage scenarios for AI-DAPT, showcasing end-to-end workflows that reflect the roles and expectations of key stakeholders, thereby aligning pipeline functions with both the user needs and the technical requirements.

Section 6 specifies the functional and non-functional technical requirements that will inform the iterative development of AI-DAPT’s Minimum Viable Product (MVP), described in **Section 7**.

Section 8 provides a detailed ethics analysis, including ethical constraints, legal considerations, and trustworthiness factors vital to AI-DAPT’s responsible deployment. For each Demonstrator, ethical and legal frameworks are defined to address specific concerns such as human-in-the-loop (HITL) considerations, regulatory compliance, and data protection.

2 AI-DAPT Research Agenda

2.1 Research Agenda

The AI-DAPT Research Agenda serves as a strategic framework to guide research and development activities throughout the project. It identifies key challenges, open research questions, and prioritized topics within the scope of AI-DAPT, focusing on advancing the lifecycle of automated data/AI pipelines. This agenda is crafted as part of *T1.4 Automated AI Pipeline Lifecycle Design & AI-DAPT Research Agenda* and spans the five main phases of the Data/AI pipeline lifecycle: data design, data nurturing, data generation, model delivery, and data-model optimization. Explainable AI (XAI) methodologies, which are expected to be applied across the entire lifecycle of AI-DAPT pipelines, are also included in the project's research agenda.

2.1.1 Methodology

The Research Agenda is developed through a structured and iterative process, ensuring alignment with the project's goals and consortium expertise. The methodology includes the following steps:

Literature Review An extensive review has been conducted on existing research, technologies, and methodologies relevant to data/AI pipelines, XAI, and science-guided AI systems. State-of-the-art practices as well as challenges-limitations in these domains have been identified, emphasizing data processing & generation for AI, hybrid science-guided AI models and adaptive AI techniques. The outcome of this extended literature review has been presented in deliverable *D1.1 - AI-DAPT Automated AI Pipeline End User Needs and Scientific Technology Radar*.

Stakeholder Consultation Domain experts, end users, and consortium members have been engaged to provide insights into key challenges and opportunities in developing AI pipelines. Feedback from demonstrator stakeholders has been incorporated in the state-of-the-art analysis, to ensure practical applicability and alignment with end-user needs.

Gap Analysis Identified gaps have been assessed in existing research, practices and tools, that the AI-DAPT project seeks to address. These gaps are further prioritized based on their impact on the scalability and robustness of data/AI pipelines, also under consideration of the fields of expertise of research groups in the consortium.

Preliminary Agenda Drafting An initial version of the agenda has been drafted in *Deliverable D1.1 - AI-DAPT Automated AI Pipeline End User Needs and Scientific Technology Radar* based on the literature review, stakeholder input, and identified gaps.

Iterative Updates The preliminary agenda has been elaborated upon and refined through iterative discussions among research groups within the consortium, leveraging their fields of expertise, and leading to the updated version presented here. In this version, a more fine-grained view is provided of open research topics to which AI-DAPT intends to contribute to. As the project progresses, the Research Agenda will be revisited and refined to incorporate findings from ongoing exploratory and experimental activities. At the same time, relevant new methods and tools that appear in the data/AI landscape will be uncovered with use of the Science and Technology Radar developed under task T1.3. This iterative approach ensures that the agenda remains dynamic, addressing emerging challenges and leveraging new opportunities.

2.1.2 Research Questions

The Research Agenda focuses on critical questions across the five phases of the Data/AI pipeline lifecycle, outlined in Section 4 of this deliverable: the data design, the data nurturing, the data generation, the model delivery and the ever-ongoing data-model optimization phase (see Section 4.1: Main phases in Data/AI Pipeline lifecycle), including also Explainable AI which cross-cuts across the entire lifecycle. AI-DAPT aims to contribute to answering the following research questions within these phases:

Data Design

- How can data be effectively identified, collected, and structured to meet the requirements of AI models?

Data Nurturing

- How can AI/ML-based techniques improve data annotation, cleaning, and preparation to enhance data quality and reliability?
- What tools and methods can automate data curation while maintaining transparency and fairness?

Data Generation

- How can synthetic data generation address the challenges of limited or inaccessible real-world data?
- What measures ensure the utility and fairness of synthetic data compared to real-world datasets?

Model Delivery

- How can hybrid science-guided AI models improve the accuracy and robustness of predictions in industrial and real-world settings?
- What strategies optimize model training, evaluation, and deployment processes?

Data-Model Optimization

- What methodologies ensure comprehensive bias assessment and correction during the model training and delivery phase?
- How can data and model observability techniques detect and mitigate data drift and model degradation in real time?
- What frameworks support adaptive AI techniques to continuously refine models in dynamic environments?

Explainable AI

- How can XAI techniques provide actionable insights for both technical and non-technical users at each pipeline stage?
- What are the best practices for integrating XAI methods into automated pipeline operations while ensuring fairness and trust?

The agenda aligns with the five lifecycle phases identified in Section 4, ensuring that each phase addresses its unique challenges while contributing to the overall goals of AI-DAPT. Explainable AI is integrated as a horizontal layer, enhancing interpretability, transparency, and trustworthiness across all stages.

2.2 Data Processing for AI

2.2.1 Data Harvesting

Data Harvesting can be defined as the acquisition of large volumes of data from any given source, to be used in various applications, such as decision support, business intelligence, training of Artificial Intelligence (AI) and Machine Learning (ML) models. In the context of *AI-DAPT*, this data is harvested by collecting datasets from services, devices and sensor readings, to train AI/ML models. In the training of AI and ML models, a considerable amount of data is required [1]. Data is available in diverse formats, including files, web services/APIs, and databases, with varying levels of complexity. It may be structured or unstructured, as well as historical or real-time. To effectively meet *AI-DAPT* requirements, it is essential to adopt solutions tailored to the specific nature of the data, as handling Big Data requires approaches distinct from those used for traditional relational datasets [2].

The data can be collected in bulk as batch data or fed through data streaming services. Tools such as Apache Spark [3], Apache Flink [4], and Amazon Elastic MapReduce [5] support **efficient batch data handling** by using distributed computing to run workloads in parallel. Distributing the computation across multiple machines increases scalability and reduces processing time, making these solutions effective for working with large-scale datasets [6] [7]. As part of *AI-DAPT*, we will utilize the above technologies to allow robust batch data handling, tailored to the scale and complexity of each dataset under consideration. This will minimize delays and errors when working with big datasets, aligning with the project's performance and reliability requirements.

For the **harvesting of streams of data**, streaming platforms like Apache Kafka [8], Apache Flink, Amazon Kinesis [9] and RabbitMQ [10] can be used to process the collection of data in real-time. This allows the processing of data from sensors, devices, services, for which all are relevant for *AI-DAPT*, as well as the storage of these data streams in clusters. These solutions are scalable distributed systems that consists of servers and clients. The server side runs as a cluster of one or more servers. Some of these servers as used as a storage layer and are called brokers while others continuously import and export data as event streams [11]. The clients allow the creation of distributed applications and services that read, write and process the data received from the streams.

In some situations, especially when harvesting streams of data, there is a need for an **intermediate storage** that can store the data in a tabular format, to serve as a buffer and to save the data for a later use (to save it to a database, to consume it in a service, etc.). To do this, there are some solutions that can handle Big Data, and if necessary, we will use incorporate them in *AI-DAPT*. Apache Druid [12] and MinIO AIStor [13] are examples of distributed Object and data stores that have integration with the above mentioned technologies and, due to the distributed nature of their implementation, are scalable to fit the needs of the solutions [14] [13].

2.2.2 Data Documentation

Metadata refers to the data about a dataset, functioning as its "manual" by detailing the dataset's origin, structure, and the information needed for seamless use across systems that require it [15]. This metadata can be categorized as **descriptive**, **structural**, and **administrative**:

1. **Descriptive Metadata**: Provides information about a dataset's source, including its origin and purpose.
2. **Structural Metadata**: Details the dataset's composition, such as data field descriptions and relationships between fields, encompassing data types and their real-world semantics.
3. **Administrative Metadata**: Focuses on technical aspects, like storage location, and aids in processes like data harvesting and cleaning.

Metadata serves as the bridge between raw data and its usability in AI workflows. Its presence ensures datasets are:

- Properly documented to enable seamless integration across platforms.
- Capable of supporting high-quality data processing, including cleaning and transformation.

Since data is crucial for AI, it's essential to ensure that data is both abundant and easy to work with. This is where principles like the FAIR (Findable, Accessible, Interoperable, and Reusable) principles [16] come in, advocating for data usability:

- **Findable**: Metadata ensures datasets are discoverable through unique identifiers and rich descriptions.
- **Accessible**: Data and metadata should be retrievable using standard protocols.
- **Interoperable**: Metadata fosters compatibility across systems by defining consistent data types and formats.
- **Reusable**: Comprehensive documentation enables data to be understood and reused by various stakeholders.

Some standardized tools exist that comply to the FAIR data principles, such as:

Data Catalog Vocabulary (DCAT) [17], offering standardized descriptions that simplify tasks for both data providers and consumers [18].

DCAT Application Profiles (DCAT-AP) [19], which are domain-specific extensions of DCAT, providing tailored metadata guidelines for specialized use cases.

Some emerging approaches for metadata representation exist, such as the one presented by the authors in [16], which is a YAML-based format that enhances metadata readability and system compatibility. These formats simplify:

- **Metadata Generation**: Easier creation of standardized metadata.
- **Interpretation**: Improved human and machine understanding of metadata.
- **System Usability**: Streamlined integration across diverse tools and platforms.

The research that will be conducted on data documentation as a part of AI-DAPT will primarily focus on interpreting and storing the metadata of a dataset in Linked Data format using the DCAT-AP standard. Along with this, additional attention is given to making the documentation, extension and extraction of metadata user friendly. Search functionality using metadata properties and storing a history of dataset changes will assist in providing maximum transparency regarding the metadata across the AI-DAPT platform. Further research on this topic is focused on developing methods to add automated metadata discovery to ensure consistency and quality in the stored metadata. To summarize, key research in the context of AI-DAPT will focus on developing methods to support efficient handling of metadata in DCAT-AP standard, storing a log of dataset history and automated metadata discovery.

2.2.3 Data Valuation

Data valuation is the assessment of data quality used in training ML models. This process is crucial for model performance: the higher the data quality, the more accurate and precise the model's predictions will be [20]. Data valuation methodologies enable the classification of data points and datasets based on their impact, whether positive or negative. Another key aspect of data valuation is incentivizing data sharing by providing a way to estimate a dataset's value. It also enhances model interpretability by identifying the datasets, data points, or features most influential for model training [21].

According to the literature, data valuation methods for ML are categorized in various ways. One of those categorizations is presented by the authors in [16], which divides the methods into two groups them being:

- "Shapley-value-based": Data valuation methodologies that use Shapley value [18] to classify features and entries of a dataset, based on their contribution for the training of the ML model, and evaluate their impact over the ones that affect the model output the most.
- "Non-Shapley-value-based": Data valuation methodologies that use simpler methods, instead of Shapley, such as the calculus of influence functions from robust statistics.

Even though “Non-Shapley-value-based” methods are less computational heavy than “Shapley-value-based” ones, they also provide less accurate evaluations. Another classification for the data valuation methods is presented by [21], which divides them into the strategies bellow.

1. **“Leave-One-Out (LOO) Based”**: Strategies where the model is trained systematically, excluding one feature or dataset at a time. By tracking the model's overall performance for each exclusion, it is possible to determine which features or datasets have the most significant impact. These strategies are computationally expensive due to the need for retraining the model multiple times.
2. **“Cooperative Game Theory (CGT)”**: Methods grounded in CGT concepts, such as Shapley value, least core, and Banzhaf index. These techniques aim to fairly quantify the contributions of features or datasets by evaluating all possible combinations, ensuring a balanced and theoretically sound distribution of importance. Some examples of algorithms are the DataShapley [17] and the DataBanzhaf [19].
3. **“Desiderata Based”**: Approaches inspired by Shapley value variants, designed to address specific desiderata. These methods reduce sensitivity to individual data points, offering a broader, more general perspective on the overall impact of each dataset, often with improved computational efficiency and scalability. The algorithms presented in [21] from this category are the D-Shapley, Robust Shapley, BetaShapley, and the Data Valuation Reinforced Learning (DVRL).

In the context of *AI-DAPT*, since new datasets will be collected among different domains, we will use **Shapley** values to quantify the contribution of individual data points, allowing to understand the importance of different subsets of data. Several methods as explained above can be utilized, so that depending on the context, we will identify the value of different features, individual data points, or the overall impact of a dataset. Our findings will help identify valuable data and optimize data collection strategies in the fields of our demonstrators.

Moreover, *AI-DAPT’s* data valuation will assess the quality and fairness of datasets, aiming to identify potential biases that may affect outcomes. The motivation behind this is that if data input is biased, the output is likely to be biased as well [22]. Our first step will involve **exploratory data analysis** for identifying anomalies or missing values. A core task will be to identify class imbalances, and if necessary, apply reweighting or resampling techniques. Furthermore, we will utilize state-of-the-art open-source tools like **IBM AI Fairness 360** [23]**,** which offers metrics and algorithms that detect and mitigate biases in data, helping ensure fairness across different attributes.

2.2.4 Data Annotation

Data annotation is the process of adding information to a dataset, such as labels or descriptive text, to provide AI/ML models with more context during training. This annotation enhances the identification of key data aspects, such as patterns, which can improve model accuracy and reduce time spent in pre-processing before training.

Data annotation goes beyond simple categorization; it can add rich information and context depending on the scenario and use case. Below are some data annotation types, according to the authors in [24]:

- Label classification – Used for basic categorization of the data.
- Intermediate labelling – Provides context to the data depending on the targeted use case.
- Assessment scores – Scoring the data points, much like in data valuation, regarding annotation reliability.
- Custom labelling – Usage of tailored labels to a specific use case, according to the desired output.
- Temporal Sequence Tagging – Used to capture the order of events or actions.

The approaches used in data annotation can be divided into two big groups: manual annotation and automated annotation.

Manual Annotation is the most basic way to perform data annotation. It requires an expert to go through the datasets that need to be labelled, providing domain expertise over raw data. This type of annotation is not only time consuming but is also costly and requires significant manual effort in order to guarantee high-quality annotations [25].

To tackle the challenges of the previous group, there has been an increase in the development and usage of **Automated Annotation** approaches, including both fully automated and semi-automated methods. The authors of [25] highlight techniques such as natural language processing (NLP), parsing, and machine learning for automation, often in combination with expert human annotators to further minimize annotation time. Additionally, [24] introduces the use of large language models (LLMs) as a promising strategy to enhance the accuracy of automated annotations. In [26], the authors review two semi-automated annotation systems in the context of Human Activity Recognition. These are:

- Active Learning (AL) – Integrates human feedback by prioritizing uncertain instances for annotation. This approach enables human annotators to correct errors, improving data quality while minimizing the number of instances requiring manual labelling.
- Transfer Learning (TL) – Utilizes pre-existing annotated datasets to train models for application on new datasets, reducing annotation efforts and enhancing the accuracy of algorithms for similar activities. These systems are particularly useful when no annotated data exists for a specific task, allowing knowledge transfer and generalization across datasets. However, challenges include identifying relevant datasets and fine-tuning the approach for optimal performance.

Concrete research activities around data annotation in AI-DAPT will focus on the investigation and assessment of up-to-date automation techniques which can range from simple rule-based automation through statistical methods up to AI based methods. While our data annotation service should be flexible enough to cover a wide range of data annotation needs, we will set a particular focus on time series and tabular data, which is of special interest to the project demonstrators. Along with a survey of the most relevant and up to date automated annotation techniques, an investigation of existing open-source tools and libraries will be conducted which should ultimately lead to additional performance, simplicity and reliability of our service.

2.2.5 Data Cleaning

Data quality is a significant factor when training ML models or using data for decision making. In the project, we focus on two types of data, (a) tabular and (b) time series, and on two important and common quality issues: (a) outlier/anomaly detection and (b) missing values imputation. The former refers to detecting values (individual points or entire sequences) that significantly deviate from what is considered expected behaviour. The latter refers to filling in gaps in the data (individual points or entire sequences) with plausible values. Missing values may occur in the original data or may result after removing outliers and anomalies.

In the past years, the focus has been on Deep Learning techniques to tackle these problems. However, this usually requires large amounts of training data, which are often difficult to acquire in practice. Moreover, it results on models that are tailored to a specific task and cannot be used to perform other tasks or generalize poorly when applied to data from other domains or with different characteristics. Large Language Models (LLMs) have demonstrated the ability to overcome these challenges, by being pre-trained on very large volumes of data in a self-supervised manner and then being applied on various tasks with little or no adaptation [27].

In the context of AI-DAPT, a research topic we intend to investigate is how to leverage LLMs to perform data cleaning on tabular data [28] [29] and time series [30]. Specifically, for tabular data, an interesting direction is to use LLMs to perform semantic table interpretation, potentially employing tailored models such as Table-GPT [31], and then utilize this information to identify outlying records or values. Another novel approach is to leverage Retrieval Augmented Generation (RAG) techniques to perform missing values imputation [32]. In the case of time series, recent works have started to explore the potential of LLMs in performing various tasks, including anomaly detection and imputation [33]. An intriguing research question in this case is how the use of an LLM could enable the ability to incorporate domain-specific knowledge when performing these tasks.

2.2.6 Semantic Reconciliation

In the training of AI and ML models, a high volume of data is required [1]. To fulfil this need, it is often the case that several different datasets from different sources need to be used. These datasets are, in most cases, heterogeneous and can be structured or unstructured data [34] [2]. It is important that these datasets are homogeneous, that is, that they conform to a common schema between them, as there is a need to have common features between different training datasets to allow the use in the training of models. This presents a challenge, as the large volumes of data at play, when working with Big Data, the complexity of the data and, in some cases, the lack of standardization of data, make can it very difficult to harmonize these datasets to have a single common schema [35].

In recent years, a few studies focused on tackling this challenge by presenting different approaches to semantic reconciliation frameworks [36] [37] [38] [39]. Some have integrated **information fusion** techniques to reconcile the data into a single representation [37] [38], others use an **ontology-based framework** to an integration of heterogeneous data sources [36], while in other cases they have used **Semantic Web** technologies [39].

To tackle this challenge, AI-DAPT will work on the knowledge and approaches implemented by UNINOVA and SUITE5 in previous projects [40] [41], whereas **data harmonization service** work on conducting semantics and syntax harmonization of heterogeneous data models into common schemas. The harmonization service is in the position to aggregate the different data representations and models into a global schema which we call the Harmonization Schema. This will be the basis for the semantic annotation of data and will contain the information on the structure, data types and units of measurement that the data should have after the semantic annotation process of the raw data is completed. that will be the basis for the semantic annotation of data and contains the information on the structure, data types and units of measurement that the data should have after the semantic annotation process of the raw data is completed.

The **Mapping Schema** is the schema that the tool uses to make the connection between the raw data fields and the Harmonization Schema. The harmonization service reconciles measurement units and data type into common units and types for all the datasets, to guarantee the interoperability of the data without the need of further pre-processing.

While the Harmonization Schema is a pre-defined singular, fixed schema used for a given type of data to define the semantics and syntax model of the data, the Mapping Schema changes with the data sources used. This means that when working with a new data source, there is a need to create a matching Mapping Schema. The Data Harmonization Service provides an **autonomous Mapping Schema creation** by vectoring the entities of the raw data schema and the Harmonization Schema, and performing Cosine Similarity to measure the similarity between the entities of both schemas. Using the measured similarities, the tool creates a Mapping Schema that matches the most similar entities of both schemas between each other automatically.

2.2.7 Data Feature Engineering

Feature engineering is a process that aims to enhance model performance by refining the dataset's variables. This involves two key strategies:

- Feature Selection: This entails strategically choosing the most valuable features. Benefits include mitigating overfitting, improving model interpretability and efficiency, and addressing dataset challenges like multicollinearity.
- Feature Extraction: This involves combining and transforming existing features to create a new feature space. This transformation simplifies complex and non-linear relationships within the model's features, making them easier for training algorithms to manage.

Feature selection methods are techniques used to choose the most relevant variables or features in a dataset for use in machine learning models. These methods can improve model performance, reduce computational complexity, and enhance interpretability. There are three main categories of feature selection methods: filter methods, wrapper methods, and embedded methods. In any case, the choice of feature selection method depends on factors like the dataset size, the number of features, the type of features, and the computational resources available. Filter methods are generally faster and simpler but may not be as effective as wrapper or embedded methods. Wrapper methods offer better performance but are more computationally expensive. Embedded methods provide a balance between performance and computational cost. Understanding the characteristics of each method is crucial for selecting the most appropriate approach for a specific machine learning task.

In the next lines, the main characteristics of these methods are presented.

Filter methods select features based on their individual characteristics and their relationship with the target variable. They are independent of the machine learning model used. Univariate filter methods evaluate each feature independently, while multivariate filter methods consider the relationships between features.

- Correlation: This method measures the linear relationship between two variables. High correlation indicates redundancy, allowing for the removal of one of the variables.
- Statistical and Ranking: These methods evaluate features based on statistical tests like mutual information, chi-square, and ANOVA. Features are ranked based on their scores, and the top-ranking features are selected.

Wrapper methods use a machine learning model to evaluate the performance of different feature subsets. They involve an iterative process of adding or removing features and assessing the model's performance. While computationally expensive, wrapper methods often lead to better model performance compared to filter methods. Examples include backward, forward, and recursive feature elimination.

Embedded methods incorporate feature selection into the model training process. They combine the advantages of filter and wrapper methods but can also be computationally intensive. These methods often involve training a model, obtaining feature importance scores, and selecting the top-ranking features.

- Lasso Regularization: This method uses L1 regularization to shrink less important feature coefficients to zero, effectively performing feature selection.
- Tree-Based Algorithms: Algorithms like Random Forest, Decision Trees, and XGBoost calculate feature importance based on the reduction in impurity.

Feature Extraction on the other hand is the process which transforms raw data into features that are better suited for machine learning models, enhancing their ability to represent the underlying problem and make accurate predictions. Feature extraction can be achieved by different methods, such as:

- Dimensionality Reduction: These methods reduce the number of features in a dataset by projecting it into a lower-dimensional space. This is done while preserving the most important information, which can help to improve model training efficiency and prevent overfitting. Popular such methods are:
- Principal Component Analysis (PCA): A widely used technique for dimensionality reduction. It maintains the maximum possible information but can sometimes sacrifice feature explainability, making it less suitable when understanding the model's decision-making process is crucial.
- Autoencoders: Neural networks that learn to recreate their input data. In the process, they generate new, powerful features in a lower-dimensional space.
- Encoding: This method converts data into a format that machine learning models can easily interpret. The three most known methods of encoding are:
    - **One-Hot Encoding:** Transforms categorical data into binary vectors, where each category is represented by a single "1" in a column of zeros.
    - **Label Encoding:** Assigns a unique numerical label to each category in a categorical variable.
    - **Ordinal Encoding:** Encodes categorical data with an inherent order, preserving the relationships between categories.
- Data Transformations: These modify data to improve its suitability for machine learning models. Examples of data transformation include:
    1. Datetime Feature Extraction: Extracts meaningful features like hour, day of the week, etc., enabling models to capture temporal patterns.
    2. Scaling and Normalization: Adjusts the numerical features to a common scale, preventing features with larger values from dominating the learning process.
    3. Mathematical Transformations: Applies functions like Box-Cox, Logarithmic, or Power transformations to improve the distribution of numerical data and enhance model performance.
- Binning: Groups numerical values into discrete bins or categories, simplifying the data and making it easier for models to learn relationships.
- Time Series Specific Techniques: Such methods analyse time-dependent data. These include:
- Autocorrelation (ACF & PACF): Measures the correlation between a time series and its lagged values, helping to identify patterns and trends.
- Seasonal-Trend Decomposition using LOESS (STL): Decomposes a time series into its seasonal, trend, and remainder components, facilitating the analysis of each component separately.
- Aggregations & Grouping: Summarizes data by calculating statistics, often grouped by specific features. Commonly use methods are rolling window aggregations, such as moving averages or cumulative sums.
- Outlier Detection: Identifies extreme values in a dataset that deviate significantly from the norm. These outliers can be interesting data points for further investigation or potential errors that need to be addressed. Common methods include the Local Outlier Factor, DBScan, and Z-score.
- Imputation: Handles missing values in a dataset by filling them in with estimated values.

Regarding the research focus of AI-DAPT in feature engineering, our aim is to deliver an efficient framework for generating highly informative features that improve the accuracy and effectiveness of machine learning models. This will provide methods that are working both on Feature Selection as well as Feature Extraction/Preprocessing, starting from the needs of the demonstrators and extending the methods portfolio to accommodate also other cases. AI/ML methods will be used in the Feature Engineering process, to automate feature engineering and provide predictions.

2.3 Synthetic Data Generation

Synthetic data generation is essential when real-world data is scarce, sensitive, or costly to obtain [42]. It is the process of creating artificial data that follows the statistical properties of real data. Organizations utilize it to train and test data-intensive models, like neural networks, without relying solely on limited real-world observations. Synthetic data can also address issues in existing datasets, such as balancing classes, simulating rare events, creating hypothetical scenarios, making models more robust and versatile [43].

Simply oversampling the minority class with replacement does not significantly improve results and may be prone to overfitting. SMOTE [43] revolutionized this approach by generating synthetic instances for the minority class. Based on the available instances in the minority class, SMOTE identifies nearest neighbours and creates new instances along the line segments joining them, thus preserving class characteristics. Furthermore, SMOTE introduces an element of randomization to ensure diversity among the synthetic samples.

Several variations of the original SMOTE algorithm were proposed in recent years. For example, Borderline-SMOTE [44] oversample only the minority examples near the borderline, focusing on areas where the class separation is less distinct. Safe-Level-SMOTE [45] prioritizes generating synthetic instances for instances that are farthest from the decision boundary, maintaining the class distribution while maximizing the margin between classes. ADASYN [46] uses a weighted distribution for the minority examples according to their level of difficulty in learning. It focuses more on harder to learn data points by creating density distributions to decide the number of synthetic samples to be generated for each minority data example. Depending on the application at hand, choosing the appropriate variation of SMOTE can lead to better results in handling class imbalance.

Synthetic data generation has been regarded as a privacy-safe alternative to real data, and therefore lately started being utilized in areas such as the health care domain [47], that are significantly important for AI-DAPT. An important work that was utilized in this domain was Synthea [48], which supports longitudinal patient record generation, and provided one million synthetic patient records freely available online. Subsequent works utilized Synthea to specifically generate Type 2 diabetes synthetic datasets. For example, Liu et al. [49] utilized those to support urban planning for improved living, with example findings being that walkability of the environment around an individual being protecting against the development of type 2 diabetes. Moreover, Kaur et al. [50] utilized Bayesian Networks to learn probabilistic graphical structures and simulated patient records, while Baowaly et al. [51] synthesized electronic health records using GANs.

Large Language Models (LLMs) are powerful tools for generating text [52]. Trained on extensive corpora, they understand patterns and relationships in written language, enabling them to produce meaningful text based on input prompts. They excel in tasks that can be framed as prompts and answered in text form. Synthetic data generation aligns with this requirement. An initial dataset can be provided to the LLM in text format along with a tailored prompt. The prompt can request novel synthetic data and, if needed, specify requirements, such as class balancing or other desired properties.

Using LLMs for synthetic data generation presents two main challenges. The first is formatting the dataset for input. Data must be converted into a text-based format, such as markdown tables, CSV, or

JSON. Additionally, because LLMs have input size limits, large [53] datasets must be segmented and processed in smaller chunks. The second challenge is that pre-trained LLMs like ChatGPT are designed for general-purpose conversational tasks. While they can generate synthetic data using prompts, they may struggle with datasets that have unique or specialized characteristics [54]. In such cases, fine-tuning the LLM for the task is necessary, but this introduces challenges with computational efficiency and ensuring accuracy [55].

In the context of the AI-DAPT research activities, we will investigate tabular data (both categorical and numerical) and timeseries data challenges in synthetic data generation. Our research will include developing efficient methods to format input datasets into suitable synthetic datasets, eliminating bias and evaluating the outputs of the methods using both metrics and visualisations. Additionally, in AI-DAPT we will investigate the generation of synthetic data from scratch, as well as the generation based on selected samples of existing real-world datasets. Furthermore, depending on the demonstrators' contents, AI-DAPT might also investigate LLM-based synthetic data generation.

2.4 Explainable AI

The growing complexity of machine learning (ML) models makes their inner workings less transparent. This limits their applicability on fields where in order to deploy an ML model in production, it is essential to understand how it operates. Consequently, the field of Explainable AI (XAI), dedicated to making opaque ML models more transparent, has garnered significant attention in recent years [56].

2.4.1 Regional Explainability Methods

XAI methods are categorized into global and local methods. Each category has distinct benefits for model interpretability [57]. Global methods provide a universal summary of the model, i.e., they explain the model's behaviour across the entire input space but may lack accuracy for specific inputs. Local methods, provide highly accurate highly accurate explanations for specific inputs, but they cannot provide a complete understanding of the model as a whole.

A new category, termed Regional Methods [58], has recently emerged to leverage the benefits of both global and local approaches. Regional explainability methods serve as an intermediary between global and local XAI, providing insights into groups of instances rather than just individual predictions (local) or the entire dataset (global).

As a novel category, regional methods face several challenges [59]. A key issue is how to partition the input space into subspaces, or in other words, how to cluster the instances into groups to create accurate explanations for each subgroup [60]. The challenge involves performing this partitioning efficiently.

Within the AI-DAPT project, we will investigate the area of regional XAI methods, focusing on addressing the challenges discussed above. We will research methods for clustering the input instances in a way that leads to accurate explanations per subgroup. Ultimately, our goal is to publish an open-source Python library dedicated to regional methods. This library will implement many current regional effect methods and will continue to be updated with new developments in the field.

2.5 Bias Detection – Fairness

Bias can be defined as the error (in the data on an AI model's outcome) that '*places privileged groups at a systematic advantage and unprivileged groups at a systematic disadvantage*' [61] [62]. Bias detection consists in methods that can identify bias in a dataset or in the operations/outputs of a system. We note that, in the same context, fairness auditing is a synonym for bias detection.

The European Union [63] defines a set of six (6) sensitive attributes: *race and ethnicity, gender, religion and belief, age, disability, and sexual orientation*. A similar set of attributes is prescribed in various anti-discrimination statutes in the United States of America [64], further including *familial/marital status* in some domains (e.g. credit scoring, advertising). Upon the selection of the sensitive attributes of interest, a series of widely used, mostly statistical, fairness measures are defined on top of them, in order to detect and quantify bias either in the data (e.g. measured on the training dataset) or in the system's outputs (e.g. measured on the test set) [65], comprising the following categories: *Independence, Conditional Independence, Separation, Sufficiency, Individual fairness* and *Counterfactual fairness* (see also D1.1 – Section 2.8 and [64]) A large part of the literature on fairness definitions and bias detection methods focuses on binary classification settings and examines bias on the outcomes of the classifier on a test set. Bias detection becomes a more challenging task, when more implicit types of bias exist, such as: (a) proxy bias attributed to correlations of sensitive attributes with seemingly non-sensitive ones; (b) subgroup (or intersectional) bias [66], where bias only appears only when considering and comparing subgroups defined by more than one sensitive attributes; (c) phenomena of gerrymandering [67], i.e. hiding subgroup bias via conveniently partitioning the subgroup space. In such cases, more elaborate bias detection methods are required; (d) bias expands beyond the test outcomes of a binary classifier to cases where bias of recourse, spatial bias, or bias of rankings/recommendations exists.

In the frame of AI-DAPT, we identify two promising research directions: (a) spatial bias detection and correction and (b) subgroup bias detection.

2.5.1 Spatial Bias Detection and Correction

Spatial bias arises when disparities in AI model outcomes or performance are related to geographic or regional differences. This type of bias often reflects underlying variations in socio-economic, cultural, or demographic factors that are unevenly represented in the data. Addressing spatial bias requires to understand how geographic disparities influence fairness outcomes, as well as methods to detect and mitigate such disparities.

Key challenges include:

- Identifying regions where systematic disparities exist in model performance or outcomes, especially when these disparities are not immediately apparent.
- Handling complex spatial structures, such as overlapping regions.
- Balancing fairness improvements across regions with the overall performance and usability of the AI system.

This direction involves exploring methods to detect and quantify spatial bias and develop frameworks that address it. The ultimate goal is to ensure fairness across geographic regions while preserving system-wide effectiveness.

2.5.2 Subgroup Bias Detection

Subgroup bias, also referred to as intersectional bias, emerges when specific combinations of sensitive attributes (e.g., gender and age) result in systematically different outcomes compared to other subpopulations. These biases are often more difficult to detect because they may only appear when analysing the intersections of multiple attributes, rather than individual ones.

Key challenges include:

- Detecting hidden biases within intersectional subgroups, which are often masked when examining attributes independently.
- Managing the complexity introduced by the exponential increase in potential subgroups as the number of sensitive attributes grows.
- Understanding how bias in smaller or underrepresented subgroups interacts with system-wide fairness metrics.

This direction aims to uncover and analyse intersectional biases to ensure that fairness considerations extend to subpopulations. The research will focus on designing robust methodologies capable of identifying disparities within subgroup structures, highlighting fairness challenges across diverse intersections of sensitive attributes. By identifying subgroup bias, the goal is to provide insights that support the development of equitable AI systems, ensuring fairness for even the most nuanced subpopulations. Specific methods and tools that will be utilized in AI-DAPT to ensure any subgroup bias detection, were also discussed in section 2.2.3, specifically discussing Data Valuation within this project.

2.6 Science-Guided AI

2.6.1 Simulator & Simulator-based Inference

A parametric simulator is a program that takes parameters as input, simulates a natural phenomenon, and produces an output. Building a simulator requires domain expertise—a scientist who thoroughly understands the phenomenon and can translate it into a series of probabilistic or deterministic steps [68].

Simulators are useful for testing scientific theories, which is difficult with black-box ML models. As white-box parametric models, simulators express scientific theories as step-by-step processes [69]. By comparing simulated data to real-world observations, we can determine parameter values that best reflect reality through a process called simulation-based inference (SBI) [70]. The set of parameter values that are probable under real world observations, can be used to validate or challenge scientific theories.

In the AI-DAPT project, we address two main challenges: modelling and inference.

Modelling requires sufficient domain knowledge to encode natural phenomena into parametric simulators. This task is complex and does not scale as each problem often needs a new simulator. Natural phenomena also include unknown steps, making algorithmic representation difficult. We will investigate these challenges by developing simulators tailored to our specific use cases where applicable.

Inference involves accurately and efficiently estimating parameter values that fit the observed data, which is often difficult due to complex simulators and the unavailability of the likelihood function [71] [72]. This complexity can lead to accuracy issues, i.e., SBI methods might produce posterior samples that don't match the true posterior, or efficiency issues, i.e., SBI methods that take too long to find posterior samples. Often, a trade-off exists between these factors [73].

In AI-DAPT, we aim to develop SBI methods that improve on both fronts: faster and more accurate inference of posterior samples.

2.6.2 Hybrid Models

In order to model the thermal building behaviour of a building and create a "digital twin", a number of options include the use of white-box, black-box or grey-box models, with the first option relying on physics principles, the second corresponding to purely data driven approaches, and the latter corresponding to "reduced order" modelling.

Resistance-Capacitance (RC) [74] models are a typical example of grey-box modelling, where a building is modelled as a network of resistances and capacitances and have been employed in many applications till now such as heat dynamics and control. On the other hand, purely data driven models / deep learning models have gained momentum due to the abundance of data and low engineering cost.

In Pilot 3 of AI DAPT project, we are interested in investigating the development of "hybrid" models to leverage the advantages of both RC and black box models. With the ability of neural networks to capture nonlinearities and correct predictions made by the RC model, our goal is to accomplish a more interpretable solution and achieve better prediction accuracy in indoor temperature forecasting and energy consumption, compared to plain black box models.

The aim is to incorporate the hybrid RC model to generate the digital twin of the core components existing within household heating scenarios (heating system, building, occupants). More specifically, the integration of hybrid RC models can assist to more accurately capture the thermal behaviour of the target building under various configurations, for instance different occupant behaviour or different heating system settings. Baseline performance estimation could directly benefit from the integration of hybrid RC models to model the building behaviour when the energy saving module of the DOMX HVAC controller is disabled. The comparison between the baseline and the energy efficient modes of operation will enable the energy efficiency improvement quantification through various KPIs (energy consumption (kWh) and cost (€) reduction, efficiency improvement (%), user comfort achievement (% of time within limits) etc.).

Finally, with the use of both a RC model and/or a "hybrid" model, we would be interested in examining the potential benefits from the use of Model Predictive Control (MPC) [75] for tuning the heating's system response when targeting multiple objectives such as thermal comfort improvement, energy consumption and cost reduction.

2.7 Adaptive AI

In AI/ML, the dynamic changes in data distribution from an evolving environment (that could be caused by user interaction, data drifting as well from changes in the operational environment itself), may lead to model performance degradation, thus reducing their accuracy and negatively impacting their use [53].

One way to respond to such issues is Adaptive AI through model re-training comes into play. There are two main retraining approaches, scheduled retraining (blind adaptation) and triggered retraining (informed adaptation). In the first scenario we define a timeline where we want to retrain our model, while this method is straightforward and easy to implement someone may lose sudden changes on the data and end up with false predictions in the meantime between retraining. Triggered model training on the other hand relies on monitoring specific metrics and setting a threshold which indicates data drift and if our metrics are below this threshold model retraining is triggered.

The main proposed practices for triggered model retraining are:

- Monitoring Multiple Metrics: By relying on different trigger metrics one can capture a broader spectrum of changes in our data.
- Asynchronous Retraining: Implementing asynchronous retraining processes prevents delays and ensures that the model remains up to date with the latest data shifts.

- Snapshot Training Dataset: It is preferred to keep training subsets which were used in case the organization wants to reproduce the training process or wants to roll back on an older version of the dataset.
- Validation Before Deployment.
- Automation: Retraining pipelines should be automated so one can eliminate human error thus making the procedure more accessible and fault tolerant.

In the context of adaptive AI for model training (retraining) in highly dynamic environments, we highlight two relevant research topics: i) using meta-learning to train global models while performing online training to adapt to changes in the incoming data, in production (i.e., inference), ii) using generative AI, as a promising candidate to predict the decision about when to trigger model retraining, by efficiently capturing and modelling complex data distributions, thus adapting to dynamic environments.

Aimed at rapidly adapting to dynamic changes (e.g., from unexpected perturbations or unseen situations) meta-reinforcement learning has shown promising results.

In [76] the authors proposed the first meta-reinforcement learning algorithm to train a global model that can rapidly adapt to changes in dynamic environments in a robotic system, using its recent experience. Another meta-reinforcement learning approach, presented in [77], is applied to novel and highly dynamic environments such as resource allocation in Open Radio Access Networks, where AI and ML are utilized to cope with highly variable demands arising in such novel networks. In addition, meta-learning approaches based on Bayes-adaptive deep reinforcement learning [78], are also found in recent literature.

A different approach to keep models updated and desired performance is to trigger model retraining. Baseline approaches are based on thresholds or periodic triggering. However, continuous monitoring of key performance metrics (for example through model observability at run/execution time) and threshold definition are still challenging, e.g., inaccurate thresholds may lead to continuous and costly, unnecessary, retraining (if thresholds are too restrictive) or to poor model performance during large periods of time, until retraining is triggered (if threshold is too relaxed).

As a solution, the authors in [79], first presented an adaptive retraining approach based on an unsupervised classifier to predict when to retrain AI/ML models, in highly dynamic scenarios.

Interestingly, in [80] the same authors propose a generative AI based approach for adaptive retraining, outperforming state-of-the-art approaches, i.e., classifier-based predictive approaches and threshold-based approaches. Specifically, the authors propose using Variational Autoencoders (VAEs), and Generative Adversarial Networks (GANs), to capture and model the distribution of data in complex and evolving environments and aimed at predicting when to trigger model retraining to ensure the desired model performance according to such new distribution of data. Synthetic data generation may also play a role here by generating data upon which models can be trained to evaluate their performance on not-seen-before data inputs and thus generate new conditions of operation, helping data scientists to decide whether these models should be considered during the deployment or the adaptation procedures of an AI pipeline.

However, retraining a model does not always solve the problem of model performance degradation, and it might be in certain cases also costly and time consuming. Although automated solutions can be very helpful, they are not foolproof. Sometimes human oversight might be necessary because the changes on the data can be too complex or totally unpredicted like the COVID pandemic. So, depending on the data drift analysis one can take different actions.

For example, if the data drift is severe one might need to completely rebuild the AI/ML model. This means there would be a need to choose a different algorithm, conduct feature selection, reweight the samples in our training data, apply domain adaptation strategies or a combination of these.

Furthermore, machine learning methods are very helpful to automate a process and make precise predictions but sometimes might fail to adapt to changes, for example one may have data drift but not have the training data for some days. In situations like this one should have a fallback plan which may consist of human expert predictions, a rule-based model which is not so precise but more fault tolerant or another type of non-ML model, for example classic statistical models.

Another solution is to combine machine learning with domain knowledge and apply business logic on top of the model where a human expert can add factors in the pipeline that he/she knows that will positively affect the model like a custom loss function or set another probability threshold that separates the classes.

In the context of the project, we aim to work on the aforementioned aspects, targeting both blind and informed model retraining, as well as model rebuilding, usage limitation and model replacement, aiming to deliver an adaptive AI framework that includes Human in the Loop but accelerates certain parts using machine intelligence. As such, we will strongly link the adaptive services to the data and model observability services of the project. On top of that, we will investigate how certain adaptation techniques can be used, based on human-generated input and rules to trigger their initialisation and the execution of automated activities that are non-intrusive to the operation of the already deployed AI pipelines. The methods that we will build, will be awaiting human validation in order to be re-deployed automatically and instantiate a new “adapted” AI pipeline.

3 Demonstrators, Data Sources & Needs

To demonstrate AI-DAPT's usefulness, the platform will be tested and validated using real-life applications from four different demonstrators. Each demonstrator focuses on a specific domain: Health, Robotics, Energy, and Manufacturing. In the following subsections, each demonstrator will be introduced, along with their current or upcoming private datasets, and a list of open datasets that match their pilot projects to be used as a baseline for AI-DAPT.

3.1 Demonstrator 1: Health - Personalized Medicine Based on Non-invasive Glucose Monitoring

Demonstrator 1 presents AI-DAPT with the opportunity to directly impact and improve the lives of people suffering from Diabetes mellitus, a metabolic condition characterised by sustained high blood sugar levels. To monitor and consequently regulate sugar levels, diabetic people currently rely on glucometers. This technology is unavoidably invasive since it requires a blood sample to perform the measurement. Photoplethysmography (PPG) is an optical technology that can detect fluctuations in blood volume, thus offering an insightful window into the blood's content. Through the integration of Machine Learning algorithms, multiple studies have demonstrated a positive correlation between glucose levels and PPG measurements [81] [82]. Whilst such studies have collected and curated datasets, privacy concerns strongly hinder their distribution and accessibility.

The first demonstrator is a collaboration between MCS and CHARITE. The SmarKo health ecosystem combined with the CHARITE patient reach and physiological expertise is pivotal to collecting high-quality PPG and blood glucose data that is fundamental for the technological advancements of this user case. The AI-DAPT framework offers an optimised environment through which the demonstrator can explore the collected data and establish meaningful relationships to extract glucose levels from PPG data. By using the AI-DAPT framework, the time taken for the demonstrator to train and retrain models is expected to be reduced by 80%.

Table 3.1.1: Demonstrator 1, Datasets Overview

Source/ Ownership	Size (PPG samples)	Participant size	Sample Length (seconds)	PPG Sampling Rate (Hz)	Additional Labels
MCS Pre-Existing Glucose Dataset	480	10	600	50	• Age • Gender • Invasive Glucose Value (every 10mins) • Systolic blood pressure (mm Hg) • Diastolic blood pressure (mm Hg)
MCS/CHARITE new dataset	600	200	600	100-200	• Age • BMI (kg/m^2)

					• Fasting glucose (mg/dL) • Systolic blood pressure (mm Hg) • Diastolic blood pressure (mm Hg) • Gender • Accelerometer (M/s^2) • Gyroscope (rotational velocity)

Table 3.1.1 outlines the two main data sources that are directly relevant to this project. Whilst there exist open-source PPG datasets, most were collected from fingertip readings, and therefore cannot be directly merged with MCS' wrist-based PPG measurements. Even so, it is still useful to keep such data as it can be useful for synthetic data generation and model generalisation purposes, especially if there is positive correlation between the wrist and fingertip blood density trends. In [81], PPG signals were collected to measure blood pressure, alongside an additional secondary data field contained the patient's illnesses. [82] used this additional field of data to train a classifier that detects diabetes based on PPG signals. Such data can be used to pre-train the feature layers of a data driven model based on a binary classification task.

The MCS data format is based on ten-minute trials. Here the PPG signals are sampled at a minimum rate of 50Hz for the entire duration of the test, resulting in sets of 500 data points that correspond to a single Glucose measurement collected at around minute five. The pre-existing MCS dataset gives a good insight to the expected tendencies of the data to be collected, as well as the data cleaning steps required. The following methods must be carried out to convert the optical raw data shown in Figure 3.1.1 to digital usable data shown in Figure 3.1.2:

- Calibration of the optical raw data / compensation of hardware differences
- Calculation against base values of standardized measurement data
- Cutting into equal windows
- Replacing dropouts
- Managing entries that might not be stored in chronological order due to communication protocols.
- Remove redundant columns (ex. additional Led sensors or sensors that are not activated)

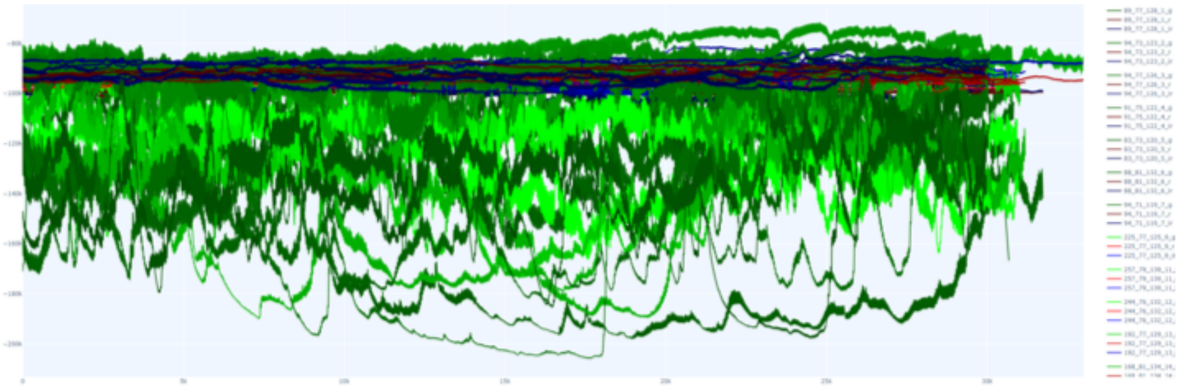


Figure 3.1.1: A visualisation of the raw PPG data where colour is chosen according to the channel and brightness is increased to for higher corresponding blood glucose values

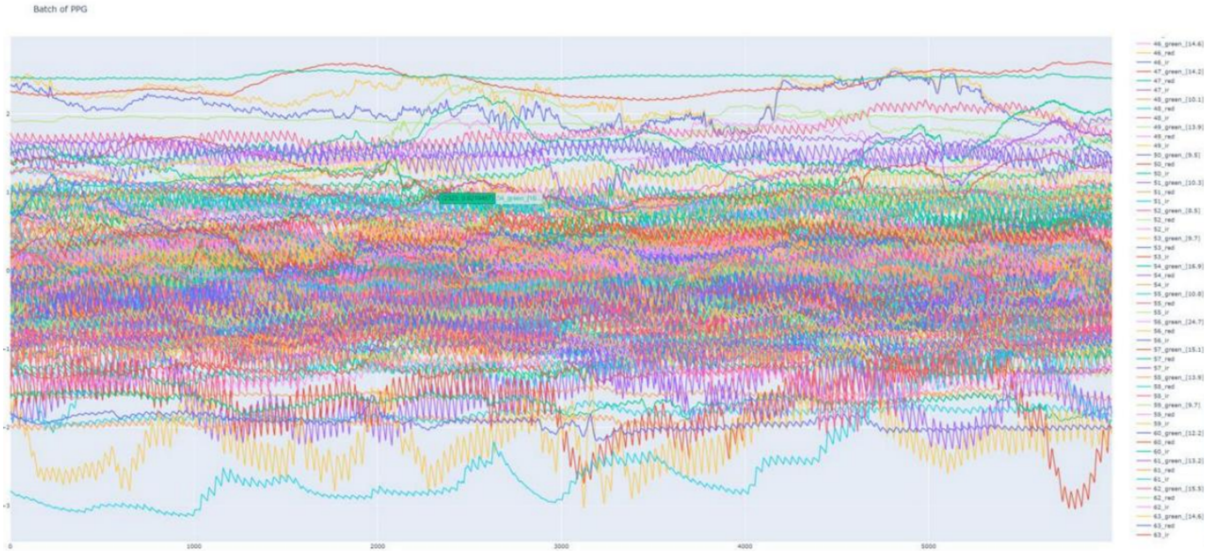


Figure 3.1.2: A visualisation of processed PPG signals

3.2 Demonstrator 2: Robotics & Cognitive Ergonomics – Human-Centred Automation

In recent years, the widespread use of wearable technology has significantly impacted our daily lives. These devices have become important sources of time-series data in the field of AI. With advancements in models that incorporate temporal dimensions, accurately analysing complex time-series data is now possible. One key area of research is detecting stress levels solely based on wearable data. Demonstrator 2, represented by IMECH and MADE, focuses on robotics and cognitive ergonomics, aiming to optimize working conditions through effective human-machine collaboration. To achieve this, they plan to monitor the working environment and the employees' emotions and mental state using wearables and then process the collected metrics and insights. Through the analysis of those insights, they want to increase worker well-being, measured through questionnaires and surveys, improve product quality, measured by the reduction in wasted components due to stress, and increase task completion speed. The insights will allow optimal identification of adjusting the behaviour of robots to minimize the stress of humans working in their environments.

Currently, the datasets are not completed yet, but they are in the process of creating them. They plan to gather data from (i) their partner companies and (ii) internally from their workers, which will consist of sensor data and real-time physiological signals from wearable devices. The data will be manually annotated based on feedback from operators and will be generated only once. More information regarding their datasets can also be found in table 3.2.1. Furthermore, there are several open datasets available for analysing and detecting stress from wearable devices which can be used as a baseline for their pilot, as listed in the table 3.2.2.

Table 3.2.1: Demonstrator 2, Datasets Overview

Dataset Name	Description	Example of Attributes	Format	Sample Rate	Source / Ownership
IMECH biosignals dataset	Dataset from partner companies, which will consist of sensor (**timeseries**) data: physiological **real-time signals from wearable devices**. The data will be **manually annotated** based on operator feedback (directly or through interviews) and, possibly, through medical experts' evaluation.	**EDA**: electrodermal activity, **range**: 0.01-20 µS, **TEMP**: body temperature, **range**: 35-42°C, **ECG**: electrocardiogram, eventually multichannel, **range**: 0.5-4 mV, RESP: respiration signal, **EEG**: electroencephalogram, eventually multichannel, **range**: 10-100 µV, **EMG**: electromyogram, **range**: 0.1-5 mV, EOG: electrooculogram, **range**: 0.05-3.5 mV, **PPG**: photoplethysmogramma, ACC: triaxial accelerometer (3 columns), **TASK**: text label for type of activity or stress / inattention level persistent data for subject baseline definition (e.g. age, weight, height, if smoker, etc.)	CSV / PARQUET & JSON	Likely in 5-15 min batches with a sampling frequency specific of the signal, in range 4-500 Hz (generated once)	IMECH
MADE Scenario dataset	Internal Dataset, which will consist of sensor (**timeseries**) data: physiological **real-time signals from wearable devices.** The data will be **manually annotated** based on operator feedback (directly or through interviews) and, possibly, through medical experts' evaluation.	**ECG**: Timeseries, Float **Oxygen Level**: Timeseries, Float **Noise Meter**: Timeseries, Float **Thermal Camera (elaborated)**: Timeseries, Float **Ambiental Sensor**: Timeseries, Float **HRV**: Timeseries, Integer Data for subject baseline definition (e.g. age (Integer), weight (Float), height (Float), if smoker (Boolean))	CSV/JSON	Likely in 15-10s batches	MADE

Table 3.2.2: Public Datasets related to the 2nd Demonstrator

Dataset	Description	Size
[83]	Data from chest and wrist-worn devices, in order to detect stress from self-report data for 15 subjects	2.5GB
[84]	Accelerometer data during 3 exam sessions (1.5h, 1.5h, 3h) and their grades of 10 subjects.	82MB
[85]	Accelerometer data during different tasks (sleep, walk, uphill walking, auditive) for 13 subjects	1.7GB
[86]	Accelerometer data throughout the day, along with stress which was reported before and after sleep for 22 subjects	23.5MB
[87]	Accelerometer data before and during the performance of mental arithmetic tasks. The data have 60sec recordings for 36 subjects	175MB
[88]	Accelerometer data during stress (spider-fearful individual watched spider clips) for 57 subjects	1.1GB

3.3 Demonstrator 3: Energy – Cross-vector Residential Demand-Response Through Smart Heating

Understanding and predicting household energy consumption and costs is essential, especially given the recent surge in energy bills and the inefficiency of the existing EU building stock. Current approaches face challenges, as most solutions like smart thermostats are limited and do not effectively manage demand or consider consumer behaviour. Accurate forecasting of energy consumption and peak-load is crucial for optimizing energy use and enhancing efficiency. For this effort, demonstrator 3 (ZENITH and DOMX), is dedicated to enhancing personalized load and price forecasts in energy consumption, along with improving the accuracy of predictions for demand response patterns. Various KPIs will be considered, such as the reduction of error for load and price forecasting and the increase of flexibility offering.

In this pilot, a plurality of datasets is considered. They have the “Simulated Temporal Power Consumption Dataset”, which as its name implies, it’s a synthetic time-series dataset generated for 2 years for 100 residential and commercial buildings. The data represent the baseline power consumption for each building using heat conduction models. This dataset contributes to their second dataset, the “Temporal Power Consumption”, which is a time-series dataset collected for 2 years from 100 real residential and commercial buildings. Moreover, they have the “Static Dataset on Pilot Population”, which contributes to the latter dataset’s measurements. Additionally, there’s the “Aggregated Historical & Geographical gas/power consumption”, where it holds large-scale gas / power consumption on district level. The “Individual Historical gas/power consumption” contains for each house specifically the historical gas/power consumption, where the “Aggregated gas/power consumption on delivery point”, includes real-time consumption for each house. Lastly, they have the “Energy Trading Spot System (ETSS) data”, and “Natural Gas Trading Platform data on Intraday Market” datasets. An overview of their datasets is provided in table 3.3.1. Additionally, table 3.3.2 contains open datasets that could also be used as a baseline for their pilot.

Table 3.3.1: Demonstrator 3, Datasets Overview

Dataset Name	Description	Example of Attributes	Format	Sample Rate	Source / Ownership
Timeseries dataset on space heating	**Time-series dataset collected for 2 years from 100 residential and commercial buildings**	Heating: Indoor/outdoor temperature, climate comfort, heating/hot water usage User: Comfort limits, user schedules and preferences, app interactions Energy: instant power, energy per usage scenario	CSV	All parameters collected at maximum rate per minute – collected once	DOMX
Simulated baseline performance dataset	**Synthetic time-series** dataset generated for **2 years** for **100 residential and commercial buildings.** (The baseline consumption is synthetically generated, under the assumption that the DOMX smart thermostat functionality is disabled.)	Energy: Baseline Daily energy per usage scenario	CSV	All parameters generated per consumer per-hour.	DOMX
Static dataset on pilot population	**Static dataset** characterizing the **pilot population** (contributing to the time-series dataset's measurements).	Building data: size, energy class, construction year, approximate location, heating zone Occupants: # of occupants, age groups, income level Contract type: fixed, dynamic, kWh price	CSV	Collected once, static	DOMX
Aggregated Historical & Geographical gas/power consumption	**Large-scale** (**aggregated**) **gas / power** demand dataset	time, day-ahead total load forecast, actual total load, RES, etc.	Excel	Hourly/Daily	ZENITH

Individual Historical gas/power consumption	**Individual historical gas / power** consumption dataset: on the **delivery** point (house)	region, municipality, post code, address, point of delivery, etc	Excel	Hourly/Daily	ZENITH
Energy Trading Spot System (ETSS) data	Information on energy trading activities, including pricing, volume, and transactions in spot markets	Trading Date, Trading Series, Contract, Start Price, Max Price, Min Price, Last Price, Closing Price, Previous Closing Price	Excel	Hourly/Daily	ZENITH

Table 3.3.2: Public Datasets related to the 3rd Demonstrator

Dataset	Description	Size
[89]	Hot water heater, dryer, and HVAC consumption, for 6 houses, along with their building specifications, occupancy, weather data and neighbouring building energy use.	16.5GB
[90]	Annual electricity, natural gas and district steam consumption, greenhouse gas emissions, energy cost, and weather-normalized data for 139 municipally operated buildings.	160.3KB
[91]	Energy Consumption, and occupancy on various University Campus Buildings in Delhi, India.	593.6 MB

3.4 Demonstrator 4: Manufacturing – Predictive Maintenance of Production Assets

The availability of production equipment is crucial in manufacturing to ensure timely and cost-effective processes. Regular maintenance and unexpected repairs must be managed without disrupting production, requiring coordination among operators and service providers. Employing Natural Language Processing (NLP) can enhance maintenance efficiency by streamlining tasks such as scheduling, tracking, and documentation, thus reducing manual effort and ensuring timely availability of spare parts and certified technicians. This leads to better management of maintenance activities, minimizing production stops and improving overall operational efficiency. For this effort, demonstrator 4 (OHS and BIBA) aims to enhance maintenance quality and efficiency in manufacturing while simultaneously detecting events to reduce costs and extending the capabilities of predictive maintenance services, ultimately resulting in reduced costs for new contracts. Their objectives include a reduction in both times required for repairs and effort for new contracts, an increase in the remaining useful life of equipment, and a decrease in edge-device power consumption. These will be achieved through correct, understandable and comfortable predictive maintenance.

This demonstrator comprises various datasets as well, including both static and dynamic ones. First of all, they have the “Assets” and “Asset Categories”, which list all of the assets that can be used in their maintenance processes along with their categories. Next, they have the “Maintenance Processes”, which includes operational data and are being kept updated during the maintenance of each asset. They also have the “Additional Efforts”, which contains information about their spare parts, unplanned repair activities, along with other maintenance activities. And, lastly, they have the “Reporting”, which are datasets that are automatically generated weekly or quarterly, to summarize the maintenance processes. Table 3.4.1 provides an overview of the latter datasets. Table 3.4.2 lists open datasets that could serve as a foundation for their pilot.

Table 3.4.1: Demonstrator 4, Datasets Overview

Dataset Name	Description	Example of Attributes	Fields with Missing Values	Format	Sample Rate	Source / Ownership
Assets (Equipment)	A dataset listing all assets that can be used in the maintenance processes	name, category, equipmentNumber, instanceNumber, sequenceNumber, manufacturerPartNumber, etc	-	CSV, Excel, & JSON	Generated once	OHS
Asset (Equipment) Categories	Dataset holding information on the categories of equipment utilized during maintenance processes	Field Name, name, description, remark, subCategory, assetsOfCategory, owningProgramme, numberOf Assets, maintenaceEffortClass, criticality, etc	activities	CSV, Excel, & JSON	Generated once	OHS
Maintenance Processes	Operational data, one record per process, updates during maintenance process	name, asset, equipmentNumber, category, repair, estimatedDelivery, location, etc	location, expectedDeliveryDate	CSV, Excel, & JSON	Daily	OHS
Additional Efforts	Spare parts, unplanned repair activities, updates with new efforts and activities therein	name, assetIdentifier, asset, creationDate, category, costs, costCenter, currencyId, etc	-	CSV, Excel, & JSON	Daily	OHS
Reporting	Spreadsheets automatically generated		-	Excel	Weekly / Quarterly	OHS

Table 3.4.2: Public Datasets related to the 4th Demonstrator

Dataset	Description	Size
[92]	The system contains 607 Non-Disclosure Agreements and a trained model checks a set of 17 hypotheses (e.g., "Some obligations of Agreement may survive termination"). Then, it identifies parts of the text that clarify whether each hypothesis is entailed by, contradicting to or not mentioned by the contract.	79.8MB
[93]	Contains legal, administrative, and contractual texts.	256GB
[94]	Includes 446 contracts with parallel lain-text section-level English summaries.	510KB

4 Pilot End-to-End Data/AI Pipeline Design

The AI-DAPT concept emphasizes a data-centric approach to Data/AI pipeline management. The project aims to automate and systematically manage data and AI processes, ensuring the creation of adaptable, scalable, and intelligent data-AI pipelines. These pipelines are designed with data-driven and model-driven steps, grouped together in five main phases, namely the data design, the data nurturing, the data generation, the model delivery and the ever-ongoing data-model optimization phase. The potential and added value of the AI-DAPT approach will be exposed through the implementation of its demonstrators. Therefore, a fundamental step in the design of data/AI pipelines under the AI-DAPT concept is the identification/definition of end-to-end pipelines that fulfil the demonstrators' needs, either in their current working scenarios (as-is) or in their future use of the AI-DAPT platform (to-be). At the conceptual level, this exercise gives the opportunity to researchers in the consortium to gain a deeper understanding of each use case scenario and the characteristics of available data sources, identify challenges and possible limitations, and brainstorm on novel approaches to tackle the problems to be solved while integrating business and operational logic in each case. At the technical level, the collection of methods and tools required for the implementation of these pipelines provides partners responsible for the design and development of platform components with a first registry of the technology stack these components will have to support.

The current section documents this procedure of end-to-end pipeline design for the AI-DAPT pilots. First, the five main phases in data/AI pipelines identified in AI-DAPT are outlined. Then, the procedure followed for step-by-step pipeline design in the demonstrators is explained. Finally, the identified pipelines are categorized into *Data Preparation, (Model) Training and Evaluation, and Deployment pipelines*, to separate those pipelines that will run (once or a few times) in the exploratory stage (Data preparation, Training and Evaluation pipelines) from the well-specified ones that will run iteratively, on a schedule or on demand in the operational stage (Deployment pipelines).

4.1 Main Phases in Data/AI Pipeline Lifecycle

AI-DAPT focuses on building automated, scalable data/AI pipelines that are data-centric and adaptable, supporting lifecycle phases like design, execution, observability, and ongoing management. These pipelines integrate both data-driven and model-driven steps, aligning business and operational logic in a sequence where each output feeds into the next step. The design enables flexibility, allowing pipelines to run on schedules, in real-time (streaming), or be triggered by events.

The AI-DAPT concept follows a structured approach, involving five main phases of pipeline operations (graphically represented in Figure 4.1.1), with distinct steps included in each phase:

I. **Data Design for AI:** Begins with researchers understanding the business problem and selecting suitable datasets, under the guidance of domain experts (*Data for AI Purposing*). An automated data mining process fetches necessary data, supporting up-to-date analysis (*Data Mining/Harvesting*). Exploratory Data Analysis (*EDA*) helps understand the data's main characteristics, supported by metadata documentation (*Data Documentation*). Data quality is assessed, ensuring suitability for AI solutions through a structured cataloguing and data valuation process (*Data Valuation*).

II. **Data Sculpting/Nurturing for AI:** Focuses on enhancing data quality and relevance through AI-driven annotation, selection, and cleaning techniques (*Data Annotation, Data Selection, Data Cleaning*). AI/ML-based techniques under human supervision are used in these steps to map the data to a selected data model, and address data errors or incompleteness. At the completion of this phase the data is representative and suitable for the AI model's needs.

III. **Data Generation for AI:** In cases where real data is insufficient, synthetic data generation supplements or replaces real data, protecting privacy and reducing bias (*Synthetic data generation*). The data utility assessment ensures synthetic data aligns well with the original data distribution (*Data Utility Assessment*).

IV. **Model Delivery for AI:** In this phase, AI models are developed (*Model Training*), evaluated (*Model Evaluation*), and deployed (*Model Deployment*). The hybrid science-guided ML model concept integrates first-principles scientific models and data-driven ML models, allowing either standalone or combined approaches to optimize predictions (*Hybrid Model Definition*). Sparse modelling is investigated at this phase, for the development of efficient AI models with reduced needs for computational resources: either at the data level, for example by dimensionality reduction of the input data, or at the model level, by enforcing sparsity on a model's weights or a neural network's activations (*Sparse Model Generation*).

V. **Data-Model Optimization for AI:** This continuous phase involves monitoring and improving models and data over time, using data and model observability to detect distribution drift (*Data Observability*), track performance, and identify retraining needs (*Model Observability*), in which case adaptive model (re)training may be required.

Explainable AI (XAI) is integrated throughout, allowing human-in-the-loop intervention and fostering transparency for decision-making. The AI-DAPT project combines automation with human oversight, improving data preparation, model development, and deployment across diverse production environments.

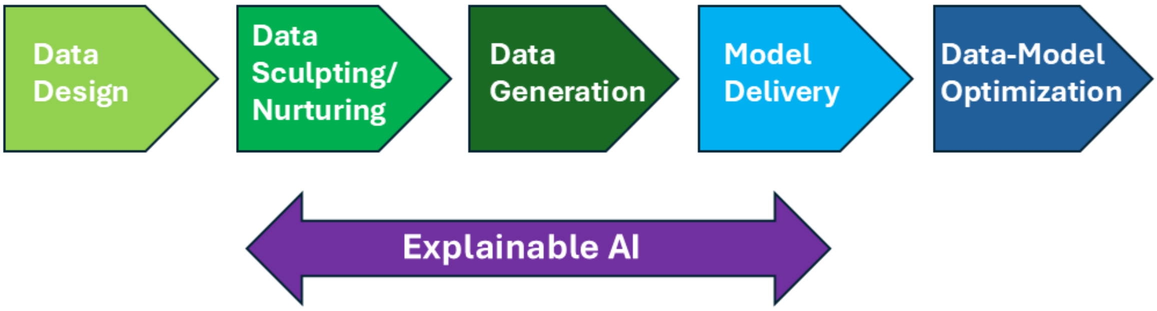
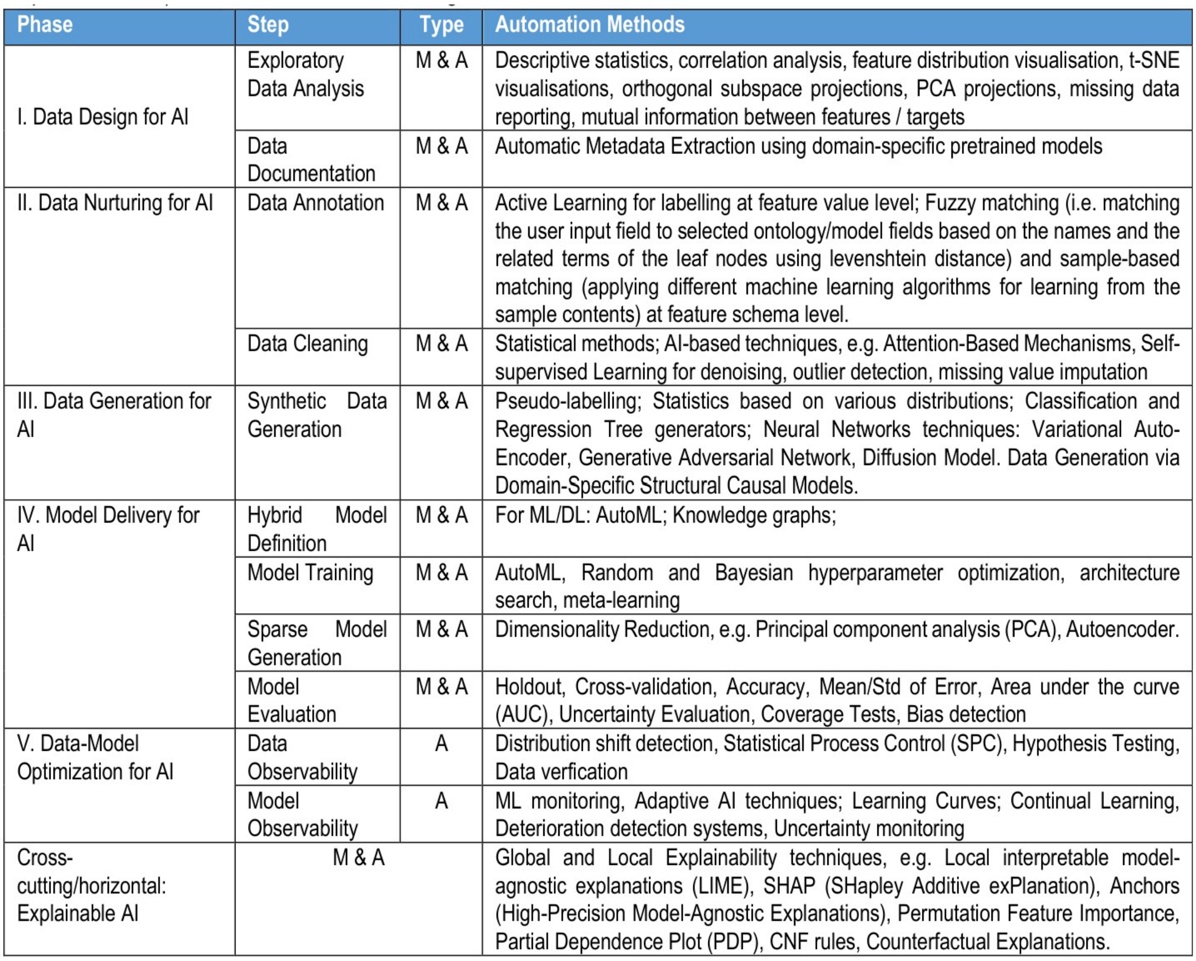


Figure 4.1.1. Main phases in AI-DAPT Data-AI Pipeline operations

Indicative methods to be applied in each phase have been identified in the project's DoA, those are presented in Figure 4.1.2 for completeness.



Phase	Step	Type	Automation Methods
I. Data Design for AI	Exploratory Data Analysis	M & A	Descriptive statistics, correlation analysis, feature distribution visualisation, t-SNE visualisations, orthogonal subspace projections, PCA projections, missing data reporting, mutual information between features / targets
	Data Documentation	M & A	Automatic Metadata Extraction using domain-specific pretrained models
II. Data Nurturing for AI	Data Annotation	M & A	Active Learning for labelling at feature value level; Fuzzy matching (i.e. matching the user input field to selected ontology/model fields based on the names and the related terms of the leaf nodes using levenshtein distance) and sample-based matching (applying different machine learning algorithms for learning from the sample contents) at feature schema level.
	Data Cleaning	M & A	Statistical methods; AI-based techniques, e.g. Attention-Based Mechanisms, Self-supervised Learning for denoising, outlier detection, missing value imputation
III. Data Generation for AI	Synthetic Data Generation	M & A	Pseudo-labelling; Statistics based on various distributions; Classification and Regression Tree generators; Neural Networks techniques: Variational Auto-Encoder, Generative Adversarial Network, Diffusion Model. Data Generation via Domain-Specific Structural Causal Models.
IV. Model Delivery for AI	Hybrid Model Definition	M & A	For ML/DL: AutoML; Knowledge graphs;
	Model Training	M & A	AutoML, Random and Bayesian hyperparameter optimization, architecture search, meta-learning
	Sparse Model Generation	M & A	Dimensionality Reduction, e.g. Principal component analysis (PCA), Autoencoder.
	Model Evaluation	M & A	Holdout, Cross-validation, Accuracy, Mean/Std of Error, Area under the curve (AUC), Uncertainty Evaluation, Coverage Tests, Bias detection
V. Data-Model Optimization for AI	Data Observability	A	Distribution shift detection, Statistical Process Control (SPC), Hypothesis Testing, Data verfication
	Model Observability	A	ML monitoring, Adaptive AI techniques; Learning Curves; Continual Learning, Deterioration detection systems, Uncertainty monitoring
Cross-cutting/horizontal: Explainable AI	M & A		Global and Local Explainability techniques, e.g. Local interpretable model-agnostic explanations (LIME), SHAP (SHapley Additive exPlanation), Anchors (High-Precision Model-Agnostic Explanations), Permutation Feature Importance, Partial Dependence Plot (PDP), CNF rules, Counterfactual Explanations.

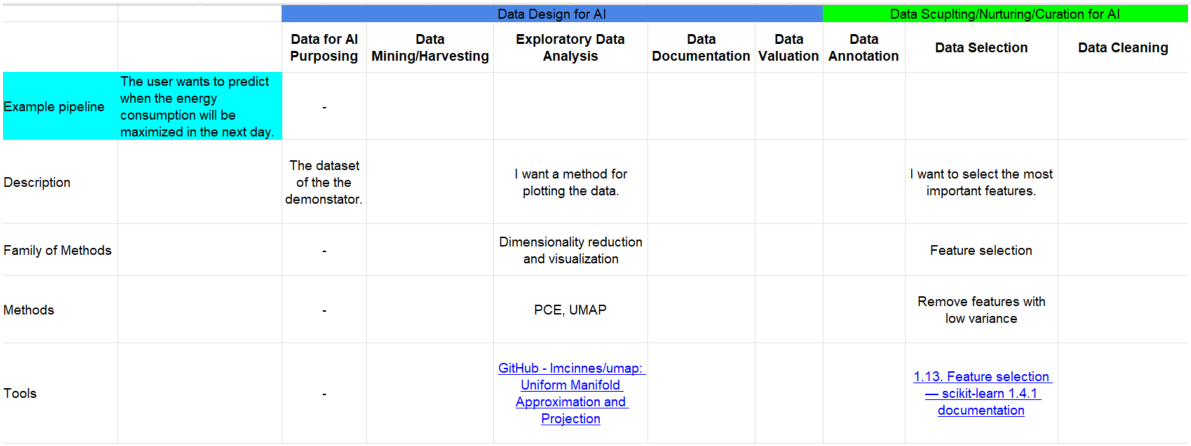
Figure 4.1.2: An indicative set of steps for the main phases of pipeline operations, as presented in the project's DoA. The type of methods in column three is annotated as manual (M) and/or automated (A).

4.2 Step-by-Step Pipeline Design in AI-DAPT Pilots

The collection of pipeline requirements, methods and tools from the demonstrators was initiated with the circulation of an Excel spreadsheet, available at the project's internal SharePoint. A separate sheet for each demonstrator was created therein, with columns annotated following the five main phases in Data/AI pipelines and corresponding steps within each phase, as described in the previous section. Demonstrator and technical support partners were then invited to construct exemplary pipelines for their use cases, indicating the needs identified and the tasks to be performed within each step of each phase of the pipeline, along with corresponding specific methods and software tools for each task/step. An example pipeline with distinct steps was also provided in the spreadsheet, to assist partners in this process, based on the following format:

- **<Pipeline X>, [short description of the pipeline]**: A name/number and a short description of the purpose of the pipeline.
- **Description**: A short description of the task to be performed within a given step of the pipeline.
- **Family of methods**: The category of methods that are proposed to perform the given task.
- **Methods**: Specific method(s) within the proposed category that seem more suitable/are selected by the pilot for the given task.
- **Tools**: Available software tool(s) implementing the proposed method(s), i.e. the URLs where the tools can be accessed.

A screenshot assisting the reader to visualize the structure of the spreadsheet is provided in Figure 4.2.1 below.



		Data Design for AI					Data Scuplting/Nurturing/Curation for AI		
		Data for AI Purposing	**Data Mining/Harvesting**	**Exploratory Data Analysis**	**Data Documentation**	**Data Valuation**	**Data Annotation**	**Data Selection**	**Data Cleaning**
Example pipeline	The user wants to predict when the energy consumption will be maximized in the next day.	-							
Description		The dataset of the the demonstator.		I want a method for plotting the data.				I want to select the most important features.	
Family of Methods		-		Dimensionality reduction and visualization				Feature selection	
Methods		-		PCE, UMAP				Remove features with low variance	
Tools		-		GitHub - lmcinnes/umap: Uniform Manifold Approximation and Projection				1.13. Feature selection — scikit-learn 1.4.1 documentation	

Figure 4.2.1: Indicative screenshot of the spreadsheet to collect pipeline needs, methods and tools from the pilots.

Initial contributions from the pilots were further discussed and elaborated with the support of WP1 partners within T1.4 activities during dedicated sessions per pilot. In cases where the pilots were already familiar with the documented pipeline operations, as part of their current (as-is) working scenarios, specific methods and tools have been proposed. There were also cases however where the pilots describe operations they would like to perform within AI-DAPT and include in their future working scenarios (to-be), so that no specific methods or tools could be proposed from the pilot's side based on experience. This is particularly evident in Demonstrator 4 – Manufacturing. In the latter, tech support partners of the pilots and WP1 partners will investigate and propose such methods and tools that could be suitable for each of these tasks.

Follow up discussions on the collected inputs led to the conclusion that most of the registered pipelines were too condensed, in the sense that these contained the entire circle of operations to address a given use case scenario, including the exploration/experimentation stage on both data and AI models, as well as the operational/delivery stage. These pipelines are therefore broken down into Data Preparation, (Model) Training and Evaluation, and Deployment counterparts. Overall, to avoid repetition of needs, methods and tools, and facilitate the reading of results from both the technical and research perspective, the proposed methods and software tools are presented collectively for all pilots in Table 4.2.1, structured according to the five main phases and steps in each phase. Pipeline steps for which the exact methods and tools to be applied are unclear, are indicated as "To Be Explored" in the table. Then, the identified Data Preparation, Training and Evaluation, and Deployment pipelines are presented in higher level in section 4.3 for each of the demonstrators, providing an overview of all types of pipelines that will be supported by the AI-DAPT framework.

Phase	Step	Family of methods	Methods	Tools	Demo
I. Data Design for AI	Exploratory Data Analysis	Descriptive statistics	Average, std, max, min, Percentiles, frequency of appearance	Pandas [95]	1, 3, 4
		Correlation analysis	Linear correlation	Pandas Dataframe Correlation method [96]	2, 3

		Dimensionality reduction	PCA, UMAP, LDA	Scikit Learn PCA Method [97] [97] Uniform Manifold Approximation and Projection [98] Scikit Learn Linear Discriminant Analysis Method [99]	1, 2, 3
		Visualization	Pairwise plots, correlation heatmaps	Matplotlib [100] Seaborn [101]	1, 2, 3
	Data Documentation	Metadata generation	Metadata based on plausibility driven by business logic	To Be Explored	4
	Data Valuation	Injection of data quality based on data documentation (alignment with business logic)	Traffic light indicator	To Be Explored	4
II. Data Nurturing for AI	Data Annotation	Label data based on Data Valuation (traffic light indicator)	Label data based on Data Valuation (traffic light indicator)	To Be Explored	4
	Data Selection	Subset selection	Selection of invalid data points	To Be Explored	4
		Feature selection	Remove features with low variance	Scikit learn Feature selection method [102]	1, 2, 3
		Time series analysis	Feature extraction, temporal correlation analysis	Pandas Dataframe Correlation method [96] PyWavelets [103]	2
		Statistical analysis	Statistics and tests	Statsmodels: Statistics and Tests methods [104]	2
		Signal decomposition	Fourier Transform, Empirical Mode Decomposition, Wavelet Transforms	SciPy: Discrete Fourier transforms methods [105] PyEMD EMD method [106] PyWavelets: Wavelet Transforms methods [107]	1, 2

	Data cleaning	Frequency filtering, resampling, missing value management	Frequency filtering, resampling, missing value management	pys: A Python Package for Time Series Classification [108]	2
		Feature engineering	Physiological signal processing, imaging	NeuroKit2: Neurophysiological Data Analysis [109] MNE: Open-source Python package for exploring, visualizing, and analyzing human neurophysiological data [110]	2
		Interpolation, imputation	Interpolation, imputation	Pandas [95]	3
		Imputation of invalid data points with synthetic (plausibility-business logic)	Imputation of invalid data points with synthetic (plausibility-business logic)	To Be Explored	4
III. Data Generation for AI	Synthetic data generation	Synthetic Time series generation	Time series GANs	Time-series Generative Adversarial Networks (TimeGAN) [111] TimeGAN: A pytorch implementation of Time-series Generative Adversarial Networks [112]	1, 2
		Synthetic TS generation	Pretrained Foundation model for Synthetic TS	Nixtla: Open Source Time Series Ecosystem [113]	2
		Personalized glucose TS generation	GluGAN: Generating Personalized Glucose Time Series Using Generative Adversarial Networks	[114]	1
		Synthetic data to replace invalid data points (plausibility-business logic)	Synthetic data to replace invalid data points (plausibility-business logic)	To Be Explored	4

IV. Model Delivery for AI	Hybrid Model Definition	Hybrid RC-Neural network	Deep Neural Network + xRyC thermal model	PyTorch [115]	3
	Model Training	ML models	Classification, regression, anomaly detection	Scikit Learn [116]	2, 3
		ML/DL models for TS	classification, regression, anomaly detection	Darts: Time Series Made Easy in Python [117] pys: A Python Package for Time Series Classification [108]	2
		DL models	LSTM, GRU	PyTorch [115] TensorFlow Keras [118]	1, 2, 3
		Foundation models	Transformers	Hugging Face [119]	2
		Gradient descent optimization methods	ADAM, AdamW optimizers	Keras Optimizers [120]	3
		Gradient boosting methods	LightGBM, XGBoost, CatBoost	LightGBM Python Package [121]	1, 3
		ML, DL, statistical models	To Be Explored		4
	Model Evaluation	Evaluation metrics	RMSE, error STD, WAPE, MAE, Pearson's r, MAPE	Scikit Learn Metrics [122] Keras Metrics [123] PyTorch Metrics [124]	1, 2, 3
		Training logs, validation, testing	Training logs, validation, testing	MLFlow [125]	2
V. Data / Model Optimization	Data Observability	Data drift detection	Monitor data drift / distribution changes	To Be Explored	1, 2, 3, 4
		Monitor Volatility	Volatility alert mechanism	To Be Explored	3
		Detection of invalid data subsets	Notification about invalid data points, avoid trends to unrealistic changes	To Be Explored	4

	Model Observability	Training/serving skew detection	Monitor model performance, assess whether re-training is needed	To Be Explored	1, 2, 3, 4
XAI		Investigate which input features are the most influential and how these affect the trained model's predictions.		To Be Explored	1, 2, 3, 4

4.3 Pipelines in AI-DAPT pilots

Typically, core data pipelines include phases I, II and III for the preparation of the training data for the AI models, while phases I, II and V are included in the case of constantly collecting the latest data. Examples of core AI pipelines include those to train an AI model (phases I-IV), or to perform inference on new data using a trained model (phases I, II, IV, V or just IV, V if an independent data pipeline is in place). In the project's operational stage, all phases may be included in combined Data/AI pipelines that can be executed on a schedule or be triggered by events.

The pipelines identified so far to address the needs of the project's demonstrators are presented in this section, according to the following categorization:

- **Data Preparation pipelines** involve phases *I. Data Design* – to harvest the raw data, investigate their properties and assess their quality, and *II. Data Nurturing* for AI – to preprocess the data, rectify errors, inconsistencies and/or missing values and engineer informative features. Phase *III. Data Generation* is also relevant in cases where the available data is scarce and Synthetic data generation is needed.

- **Training and Evaluation pipelines** include the operations necessary to train and validate a (hybrid) AI model: here we assume that the training data have already been prepared in an independent Data Preparation pipeline. Therefore, a typical Training and Evaluation pipeline includes phase *I. Data Design* – partially, to ingest the training data, and phase *IV. Model Delivery* - to experiment with different models and assess their performance. Once satisfactory performance is achieved, additional *XAI pipelines* can be configured to explain the effect of input features on the model's predictions and extract meaningful insights for the demonstrators.

- **Deployment pipelines** are combined data/AI workflows that are well established through the exploratory and experimentation stage and therefore suitable for scheduled or event-triggered ML operations (MLOps) in the production environments of the demonstrators. Examples include delivering predictions based on new data samples (phases I, IV, V), or regularly collecting new data and appending those in the processed datasets for model (re)training, in which case data and model drifts must be monitored and accounted for (phases I, II, IV, V).

Pipeline identifiers are generated accordingly for each demonstrator in the following format:

- Dx_DP_y corresponds to Data Pipeline y of demo number x.
- Dx_TE_y corresponds to Training & Evaluation pipeline y of demo number x.
- Dx_MLOps_y corresponds to Deployment pipeline y of demo number x.

An exception is made for demonstrator 3 – Energy, where three use case scenarios are examined, so in this case an additional index points to the respective scenario, for example Dx_z_DP_y points to Data Pipeline y of scenario z for demo number x (z=1, 2, 3).

4.3.1 Demonstrator 1: Health - Personalized Medicine Based on Non-invasive Glucose Monitoring

Data Preparation Pipelines

D1_DP_1 The user wants to collect and process PPG signals and medical data to construct an AI-ready training dataset for blood glucose monitoring.

I. Data Design

- Data for AI purposing: PPG signals & medical data/co factors e.g. age, weight (MCS & CHARITE datasets).
- Data Mining/Harvesting: Harvest data from Influx DB.
- EDA: Correlation analysis between measured and baseline signals, pairwise plots, correlation heatmap. Frequency domain analysis and input PPG-sample-length to glucose measurement ablation study.
- Data documentation: Create metadata based on data profiling.

II. Data Nurturing

- Data Selection: Feature selection (remove features with low variance).
- Data Cleaning: Filter out noise/movement of the subject, impute missing values.

III. Data Generation

- Synthetic data generation: Create synthetic data for under-represented target groups. The majority of subjects are diabetic, more data from healthy/younger subjects are needed.

Training & Evaluation Pipelines

D1_TE_1 The user wants to experiment with different AI models to find the optimal for blood glucose monitoring.

I. Data Design

- Data for AI purposing: Processed data from D1_DP_1.
- Data Mining/Harvesting: Harvest data from Influx DB.

IV. Model Delivery

- Hybrid model definition: The use of hybrid modelling will be explored, potentially leading to the development of hybrid models that integrate domain-specific (medical) knowledge in defining sample data weights, baselines or anomaly thresholds.
- Model Training: Train time series classification/regression models. LSTM networks and gradient boosting frameworks have been used so far.
- Model Evaluation: Evaluate model performances and compare different models with suitable metrics. Aim to improve the performance of LSTM used so far.

XAI

- Model Explainability: Identify the most important input features for the best performing model and investigate how the different values (e.g. high, low) of these features affect the model's predictions.

Deployment Pipelines

D1_MLOps_1 The user wants to predict blood glucose values for a patient based on PPG signals and co-factors.

I. Data Design

- Data for AI purposing: PPG signals collected from wearable sensor device. Co-factors retrieved from database.
- Data Mining/Harvesting: Harvest the data from wearable and database.

II. Data Nurturing

- Data annotation: Map data to domain-specific data model
- Data cleaning: Filter out noise/movement of the subject, impute missing values.
- Data selection: Feature selection.

IV. Model Delivery

- Model Deployment: The trained model is retrieved and deployed to predict blood glucose levels.

V. Data/model optimization

- Data observability: Monitor data drift/distribution changes with new subjects.
- Model observability: Monitor model performance, check whether retraining is needed.

4.3.2 Demonstrator 2: Robotics & Cognitive Ergonomics – Human-Centred Automation

Data Preparation Pipelines

D2_DP_1 The user wants to collect and process physiological signals and persistent data (e.g. age) to construct an AI-ready training dataset for inattention/stress condition evaluation.

I. Data Design

- Data for AI purposing: Wearables data (biosignals) collected as CSV/Parquet files and labelled/temporally aligned. Persistent data collected as metadata.
- Data Mining/Harvesting: Harvest data from database.
- EDA: Dimensionality reduction and visualization, time-series specific correlation analysis.

II. Data Nurturing

- Data Selection: Feature selection (remove features with low variance), correlation analysis (remove correlated data), time series analysis and extraction of useful components, (physiological) signal decomposition, feature engineering, imaging.
- Data Cleaning: Resampling, missing value treatment, frequency filtering.

III. Data Generation

- Synthetic data generation: Make data distribution uniform by generating synthetic data for under-sampled categories.
- Data utility assessment: Ensure realistic time series and persistent data generation.

Training & Evaluation Pipelines

D2_TE_1 The user wants to experiment with different AI models to find the optimal for inattention/stress condition evaluation.

I. Data Design

- Data for AI purposing: Processed data from D2_DP_1.
- Data Mining/Harvesting: Harvest data from database.

IV. Model Delivery

- Hybrid Model Definition: The use of hybrid modelling will be explored, potentially leading to the development of hybrid models that integrate domain-specific (medical) knowledge in defining sample data weights, baselines or anomaly thresholds.
- Model Training: Train time series classification and anomaly detection models, along with exploratory use of hybrid models. Both ML and DL approaches are considered, also pre-trained DL architectures (foundation models).
- Model Evaluation: Evaluate model performances and compare different models with suitable metrics. Training logs comparison, validation, testing.

XAI

- Model Explainability: Identify the most important input features for the best performing model and investigate how the different values (e.g. high, low) of these features affect the model’s predictions.

Deployment Pipelines

D2_MLOps_1 The operator wants to evaluate their inattention/stress condition based on physiological signals. – Alternatively, the shift manager wants to evaluate the inattention/stress condition of the operators based on physiological signals. This is under discussion, depending on demo partners’ decision on how the results/alerts will be presented: either directly to the operator, or collectively for all operators in a dashboard supervised by the production manager.

I. Data Design

- Data for AI purposing: Physiological signals collected from wearables.
- Data Mining/Harvesting: Harvest the data from different sources/sensors and automatically perform the temporal alignment and labelling.

II. Data Nurturing

- Data annotation: Map data to domain-specific data model
- Data cleaning: Resampling, missing value treatment, frequency filtering.
- Data selection: time series analysis and extraction of useful components, (physiological) signal decomposition, feature engineering, imaging, feature selection.

IV. Model Delivery

- Model Deployment: The trained model is retrieved and deployed to predict inattention/stress levels.

V. Data/model optimization

- Data observability: Monitor data drift/distribution changes with new subjects.
- Model observability: Monitor model performance, check whether retraining is needed.

4.3.3 Demonstrator 3: Energy – Cross-vector Residential Demand-Response Through Smart Heating

4.3.3.1 Forecasting of wholesale electricity market prices

Data Preparation pipelines

D3_1_DP_1 The pipeline user wants to collect and process TSO/ENEX data to construct an AI-ready dataset for the prediction of electricity market prices, such as the Day-Ahead Market (DAM) price.

I. Data Design

- Data for AI purposing: ZENITH Dataset: dataset that combines public data from TSO/ENEX. The dataset is in tabular form, i.e. CSV format.
- Data Mining/Harvesting: Harvest data from database.
- EDA: Pairwise plots - plot the correlation between each feature and the output.
- Data Valuation: Set a custom measure that evaluates the quality of the data, based for example on the output of correlation analysis.

II. Data Nurturing

- Data Selection: Feature selection, based on correlation matrix with pairs: feature (x_i) - output (y).
- Data Cleaning: Resampling, missing value treatment, frequency filtering.

Training & Evaluation Pipelines

D3_1_TE_1 The pipeline user wants to experiment with different AI models to find the optimal for the prediction of electricity market prices, such as the DAM price. To this aim, a training dataset is used as input from the previous pipeline (D3_1_DP_1).

I. Data Design

- Data for AI purposing: Processed data from D3_1_DP_1.
- Data Mining/Harvesting: Harvest data from database.

IV. Model Delivery

- Hybrid Model Definition: Machine Learning model, for example XGBoost, taking also into consideration hybrid domain-specific feature engineering techniques (e.g., temporal features, energy market structural features, etc.).
- Model Training: Train the model using a gradient descent technique, such as ADAM.
- Model Evaluation: Evaluate the error to the prediction (e.g. in terms of mean absolute error, MAE).

XAI

- Model Explainability: Identify the most important input features for the best performing model and investigate how the different values (e.g. high, low) of these features affect the model's predictions.

Deployment Pipelines

D3_1_MLOps_1 The pipeline user wants to regularly (every day) forecast the electricity market prices, such as the DAM price.

I. Data Design

- Data for AI purposing: The latest TSO/ENEX data.
- Data Mining/Harvesting: Harvest data from web API.

II. Data Nurturing

- Data annotation: Map data to domain-specific data model (SAREF4ENER).
- Data selection: Perform standard aggregations/resampling. Feature selection.

IV. Model Delivery

- Model Deployment: The trained model is retrieved and deployed to predict the electricity market prices, for example the DAM price.

V. Data/model optimization

- Data observability: Monitor data drift. Generate a Volatility Alert Mechanism based on Volatility measure, monitor Volatility.
- Model observability: Monitor model performance, check whether retraining is needed. Check for imports/exports in the data.

4.3.3.2 Estimation of energy savings achieved through smart heating

Data Preparation Pipelines

D3_2_DP_1 The pipeline user wants to process household data to construct an AI-ready training dataset for the estimation of the energy savings achieved through smart heating versus the baseline mode.

I. Data Design

- Data for AI purposing: Data from DOMX HVAC controller and smart phone app, DSO smart meter. - Actual consumption calculated based on heating data and user actions. Synthetic data generated as the baseline consumption, static data on households.
- Data Mining/Harvesting: Harvest the data from the different databases and synchronize them.
- Data Valuation: Comparison against consumption data collected through the DSO Smart Meter.

II. Data Nurturing

- Data Selection: Feature selection.
- Data Cleaning: Detecting device disconnections. Missing value identification and handling (interpolation, imputation).

Training & Evaluation Pipelines

D3_2_TE_1 The pipeline user wants to experiment with different AI models to find the optimal for the estimation of the energy savings achieved through smart heating.

I. Data Design

- Data for AI purposing: Processed data from D3_2_DP_1.
- Data Mining/Harvesting: Harvest data from database.

II. Data nurturing

- Data selection: Train on data from baseline days. Test and compare on days with adaptive data.

IV. Model Delivery

- Hybrid Model Definition: Hybrid Resistance-Capacitance (RC) model -Neural network. xRyC thermal model + Deep Neural Network.
- Model Training: Train on data from baseline days. Test and compare on days with adaptive data.
- Model Evaluation: Evaluate the model on test data by means of forecast accuracy metrics (WAPE, MAE).

XAI

- Model Explainability: Identify the most important input features for the best performing model and investigate how the different values (e.g. high, low) of these features affect the model’s predictions.

Deployment Pipelines

D3_2_MLOps_1 The pipeline user wants to regularly (every month) calculate the achieved energy savings per participant customer.

I. Data Design

- Data for AI purposing: Processed data from D3_2_DP_1.

- Data Mining/Harvesting: Harvest data from database.

IV. Model Delivery

- Model Deployment: The trained model is retrieved and deployed to predict the energy savings per participating customer for the current month.

V. Data/model optimization

- Model observability: Monitor model performance, check whether retraining is needed.

4.3.3.3 Explicit Demand Management of Heat Pumps for load shifting based on Day-Ahead Market Price Integration

Data Preparation Pipelines

D3_3_DP_1 The pipeline user wants to collect and process TSO/ENEX data as well as data sets coming from the heat-pump operation and user temperature profile.

I. Data Design

- Data for AI purposing: Day-Ahead Market (DAM) electricity price signals, Historical heat pump operational data, Building thermal characteristics, Indoor temperature profiles.
- Data Mining/Harvesting: Harvest data from database(s).
- EDA: Correlation matrix between price signals and heat pump performance.

II. Data Nurturing

- Data Selection: Thermal response pattern identification, thermal inertia parameters, comfort deviation metrics, load shifting potential features.
- Data Cleaning: Outlier detection and treatment, seasonal adjustment, missing value treatment, frequency filtering.

Training & Evaluation Pipelines

D3_3_TE_1 The pipeline user wants to experiment with different AI models to find the optimal integration between DAM price forecasted volatility and thermal inertia of the home.

I. Data Design

- Data for AI purposing: Processed data from D3_3_DP_1.
- Data Mining/Harvesting: Correlation matrix between price signals and heat pump performance.

IV. Model Delivery

- Model Definition: Neural Network/Machine Learning models for time-series.
- Model Training: Multi-objective optimization.
- Model Evaluation: Evaluate the price prediction accuracy, demand response effectiveness, thermal comfort maintenance, potential value generation.

XAI

- Model Explainability: Identify the most important input features for the best performing model and investigate how the different values (e.g. high, low) of these features affect the model's predictions.

Deployment Pipelines

D3_3_MLOps_1 The pipeline user wants to regularly integrate the forecasted DAM price with the heap-pump scheduled operation.

I. Data Design

- Data for AI purposing: DAM price feeds, heat pump operational status.
- Data Mining/Harvesting: Harvest data from APIs.

II. Data Nurturing

- Data selection: Continuous data quality assessment.

IV. Model Delivery

- Model Deployment: API-driven model inference.

V. Data/model optimization

- Data observability: Continuous performance tracking, data drift detection, model retraining triggers, cross-domain optimisation feedback loops.
- Model observability: Monitor prediction accuracy, demand response performance and economic impact.

4.3.4 Demonstrator 4: Manufacturing – Predictive Maintenance of Production Assets

Data Preparation Pipelines

D4_DP_1 The user wants to develop a methodology to identify and correct corrupted quantitative data related to timestamps/durations of/in maintenance processes.

I. Data Design

- Data for AI purposing: Quantitative data from maintenance activities, i.e. timestamps.
- Data Mining/Harvesting: Harvest data from PostgreSQL DB.
- EDA: Statistics on data – average values and thresholds.
- Data documentation: Format of timestamps, metadata generation based on the plausibility of maintenance activity duration – alignment with business logic.
- Data valuation: Timestamp completeness (start & end of maintenance activity). Injection of data quality as traffic light indicator.

II. Data Nurturing

- Data Annotation: Timestamp related to certain process and equipment ID. Transfer of business logic into ontology-style. Label data from previous step (traffic light indicator for data quality).

- Data Selection: Selection of data sets related to invalid timestamps.
- Data Cleaning: Correction of invalid timestamps with synthetic ones.

III. Data Generation

- Synthetic data generation: Generate synthetic timestamps to replace invalid data points.
- Data utility assessment: Mark the generated timestamps to monitor synthetic data in order to avoid drifts or biases in timestamp generation.

D4_DP_2 The user wants to develop a methodology to identify and correct corrupted qualitative data related to timestamps/durations of/in maintenance processes.

I. Data Design

- Data for AI purposing: Qualitative data from maintenance activities on equipment, i.e. condition of equipment to be maintained.
- Data Mining/Harvesting: Harvest data from PostgreSQL DB.
- EDA: Statistics on data – frequency of appearance.
- Data documentation: Data in text format representing categories, e.g. “corrosion”, “missing parts”.
- Data valuation: Plausibility (business logic).

II. Data Nurturing

- Data Cleaning: Representative qualitative data in relation to their processes and equipment.

III. Data Generation

- Synthetic data generation: Generate synthetic data to replace invalid data points.
- Data utility assessment: Mark the generated data to monitor synthetic data in order to avoid drifts or biases.

D4_DP_3 The user wants to verify quantitative data related to timestamps/durations of/in maintenance processes against related quantitative data, and vice versa.

I. Data Design

- Data for AI purposing: Corrected & labelled qualitative & quantitative data from D4_DP_1 & D4_DP_2.
- Data Mining/Harvesting: Harvest data from PostgreSQL DB.
- Data valuation: Plausibility (business logic).

II. Data Nurturing

- Data selection: Verification/cross-validation of quantitative data related to timestamps/durations of/in maintenance processes against related quantitative data, and vice versa.

D4_DP_4 The user wants to identify metadata of tools that fit criteria related to damage and high cost (investigation like "crime profiling").

I. Data Design

- Data for AI purposing: Corrected data on equipment & maintenance activities from D4_DP_1, D4_DP_2, D4_DP_3.
- Data Mining/Harvesting: Harvest data from PostgreSQL DB.
- EDA: Statistics on data – average values and thresholds, frequency of appearance.

II. Data Nurturing

- Data Annotation: Identification of tools that fit criteria related to damage and high cost.
- Data Selection: Selection of data sets related to tools that fit criteria related to damage and high cost.

Training & Evaluation Pipelines

D4_TE_1 The user wants to develop a model to recommend tool handling (intra-logistics) and plan logistics on tool transport based on weather and traffic data.

I. Data Design

- Data for AI purposing: Traffic info related to route planning, weather data, sensor data from tool (item level/edge).
- Data Mining/Harvesting: Harvest data from weather & traffic web APIs, harvest data from edge device on tool (when connected).
- EDA: Probability/forecast of rain & humidity. Statistics on data – average values and thresholds (min, max).
- Data documentation: Metadata on location, distance measurement.

II. Data Nurturing

- Data Annotation: Mark start and end points of bad incidents.
- Data selection: Identification through data analysis if conditions are good for tool transport. Highlight data on “bad” incidents with traffic light indicator (red, yellow).
- Data cleaning: Ignore “green” status samples.

IV. Model Delivery

- Hybrid model definition: The possibility will be investigated to develop a hybrid model by integrating weather effect on hoisting tool condition, define some baselines and anomalies for tool handling.
- Model training: Use historic data to correlate weather and traffic conditions to tool condition and delivery time.
- Model evaluation: Use current data to evaluate the trained model.

XAI

- Model Explainability: Identify the most important input features for the best performing model and investigate how the different values (e.g. high, low) of these features affect the model’s predictions.

Deployment Pipelines

D4_MLOps_1 The user wants to identify and correct corrupted qualitative/quantitative data related to timestamps/durations of/in maintenance processes.

I. Data Design

- Data for AI purposing: Qualitative & quantitative data from maintenance activities on equipment, i.e. condition of equipment to be maintained and timestamps related to maintenance activities.
- Data Mining/Harvesting: Harvest data from PostgreSQL DB.
- EDA: Statistics on data – frequency of appearance.
- Data documentation: Qualitative data in text format representing categories, e.g. "corrosion", "missing parts". Quantitative data on timestamps related to maintenance process (start, end).
- Data valuation: Plausibility (business logic).

II. Data Nurturing

- Data Cleaning: Representative qualitative/quantitative data in relation to their processes and equipment.

V. Data/model optimization

- Data Observability: Notification if a data set is identified as invalid and about its changes. Avoid trends to unrealistic changes.

D4_MLOps_2 The user wants to recommend tool handling (intra-logistics) and plan logistics on tool transport based on weather and traffic data.

I. Data Design

- Data for AI purposing: Traffic info related to route planning, weather data, sensor data from tool (item level/edge). Training data from D4_DP_1, D4_DP_2, D4_DP_3.
- Data Mining/Harvesting: Harvest data from weather & traffic web APIs, harvest data from edge device on tool (when connected).
- EDA: Probability/forecast of rain & humidity. Statistics on data – average values and thresholds (min, max).
- Data documentation: Metadata on location, distance measurement.

II. Data Nurturing

- Data selection: Identification through data analysis if conditions are good for tool transport. Highlight data on "bad" incidents with traffic light indicator (red, yellow).
- Data cleaning: Ignore "green" status samples.

IV. Model Delivery

- Model deployment: Fetch the trained model from D4_TE_1 to decide whether conditions are good for tool transport (traffic light indicator, good conditions: "green").

V. Data/model optimization

- Data Observability: Matchmaking of edge device predictions and real (external) data.

- Model observability: Adopt model to item-level (hoisting individual requirements /characteristics).

5 AI-DAPT End-to-End Usage Scenarios

5.1 Stakeholders

The AI-DAPT ecosystem is comprised of various stakeholders, that will benefit from the streamlined AI-DAPT Data and AI offerings, either through direct interaction with the AI-DAPT Platform or indirectly, as part of the greater data/AI value chain.

Out of these stakeholders, we identify the 2 following groups that will have a direct interaction with the AI-DAPT platform:

- *Data Scientists/Analysts and AI Operators* will leverage the AI-DAPT data and AI experimentation, design, automation and monitoring mechanisms to design optimised AI pipelines and stay in control of the models and pipelines performance at production phase.
- *Industry Stakeholders (including subject matter experts and decision makers*) will be able view the outputs of the AI-DAPT Pipelines and take decisions based on the operations of their businesses.

All other stakeholder groups, as identified in the project's DoA could have access to the AI-DAPT platform or its offerings, however we envisage that the core audience, and as a consequence the main users of AI-DAPT, will come from the two groups identified above.

5.2 End-to-End Usage Scenarios

The following section introduces the core end-to-end usage scenarios (SCEs) envisioned within the AI-DAPT Platform. They illustrate the flow of actions an AI-DAPT User should perform within the platform in order to perform a specific task. In particular the list of AI DAPT usage scenarios is the following:

- **SCE-01: Data Harvesting** *(onboard data to the AI-DAPT platform from external data sources, utilising the AI-DAPT Harvesting services)*
- **SCE-02: Data Sculpting** *(apply data processing methods over data in order to bring them a format suitable further use in AI and analytics operations)*
- **SCE-03: Synthetic Data Generation** *(resolve data sparsity and privacy issues through synthetic data generation)*
- **SCE-04: Data Valuation, Observability & Optimisation** *(monitor and improve the health, quality and value aspects of data through multi-faceted metrics, alerts and optimisation techniques)*
- **SCE-05: Data Pipeline Design and Execution** *(combine the aforementioned data manipulation and monitoring services – i.e., harvesting, sculpting, synthetic data, valuation observability - in customisable Data Pipelines supporting flexible execution)*
- **SCE-06: Data & Model Interactive Experimentation** *(improved UX during data and AI services configuration along with sandboxed experimentation with custom code for data and AI operations)*
- **SCE-07: Model Explanation** *(utilise XAI techniques across the AI lifecycle to enhance pipelines' explainability wherever required: data exploration, XAI-friendly model selection and design, utilisation of explainers and inspection of XAI visualisations)*
- **SCE-08: Model Observation & Adaptation** *(stay in control of AI operations in production phase through a monitoring view and alerts. Improve model performance though adaptive AI techniques)*
- **SCE-09: Hybrid AI Pipeline Design and Execution** *(combine the aforementioned AI design, experimentation and monitoring services in flexible AI Pipelines)*

- **SCE-10: AI-DAPT Datasets/Results Consumption** *(retrieve the outputs of the AI-DAPT services and Pipelines in order to utilise them in external applications through the flexible consumption mechanisms)*
- **SCE-11 Data Models and Analytics Models Registration** *(enhance the knowledge base of AI-DAPT with additional data models and analytics models. This scenario is targeted for Admin Users)*

The AI-DAPT end-to-end usage scenarios focus on operations constituting the core added-value of the AI-DAPT Platform for end-users. Thus, the analysis sticks to the main interaction flows for clarity, while typical platform scenarios (such as user registration and management), and standard editing/management/deletion branches are left out on purpose, as they are considered standards. Each of the SCEs listed above is presented in detail in the dedicated subsections below, and includes the scenario overview, steps, workflow diagram, involved users and benefits, challenges and success criteria.

Two main Actors are foreseen to have access and directly utilise the AI-DAPT Platform:

- **Data Scientist**: Representing stakeholder group (i), the Data Scientist will join the AI-DAPT Platform in order to experiment, design, deploy and monitor AI Pipelines that will solve their research or application-related AI and analytics problems. As data are a big part of Ai engineering, the Data Scientist will also utilise at a big extent the data services of AI-DAPT, especially in order to optimize the data that will be used in the AI Pipelines. The Data Scientist holds expertise in the areas of AI, ML and analytics and can utilise to the maximum the available services (including advanced configurations and advanced experimentation). They can work as an individual entity (e.g. a researcher) or as part of a bigger team that wants to solve a specific problem (e.g. providing the data science expertise needed to create Pipelines that will feed a business application of a stakeholder in group (ii).
- **Business User**: This actor is representing stakeholder group (ii). The Business User is expected to have bigger involvement with the data manipulation and the consumption services of the AI-DAPT Platform (although advanced configurations and AI services will also be available in case they want to utilise them). Their main interest in the AI-DAPT Platform will be apart from registering data also to explore the outputs of the AI-DAPT Pipelines (i.e. consumable artefacts) and select the ones they will utilise in their business applications. The Business User is assumed to enter the AI-DAPT Platform with a limited knowledge on AI, ML and analytics (although some high-level understanding of the domains can facilitate them to configure basic data and analytics operations), and they will get assistance from their team's Data Scientists.

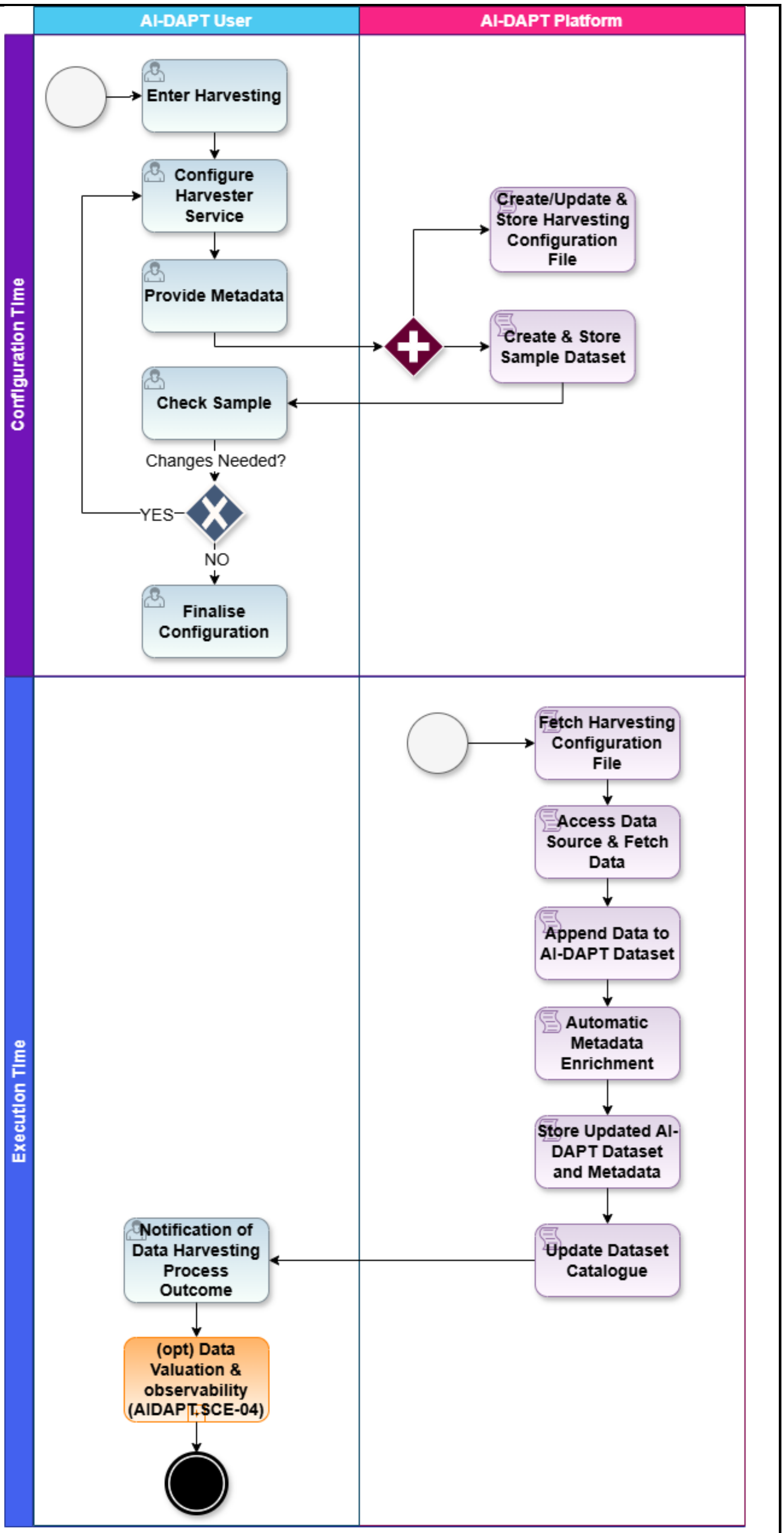
With regards to the workflow diagrams, three entities appear as interacting entities, namely: the AI-DAPT User (Data Scientist and Business User actors), the Business User Application, and the AI-DAPT Platform. Since the purpose of these diagrams is to provide a clear idea of user interactions and not on describing in detail the underlying technical communication flows and technical interactions, operations by specific AI-DAPT Services are embedded under the AI-DAPT Platform and abstracted wherever required for visual clarity.

5.2.1 SCE-01: Data Harvesting

Data Harvesting - [SCE-01]	
Scenario ID	SCE-01
Scenario Name	Data Harvesting

Scenario Overview	The authenticated AI-DAPT User wants to check-in a new dataset to the AI-DAPT Platform either as part of a Data Pipeline or through a self-standing harvesting job. The AI-DAPT User completes the harvesting setup and an initial required set of metadata that should be attached to the AI-DAPT dataset. A sample dataset is created to facilitate the setup of the next steps of the Data Pipeline (if any). The harvesting configuration is utilised at execution time in order to fetch data from the specified data source. The dedicated AI-DAPT dataset is populated with the fetched data, while additional dataset metadata are automatically extracted/generated. The populated dataset is available in the AI-DAPT Scalable Storage Services for further pre-processing (as part of a Data Pipeline), or for use in AI Pipelines. Optionally, the AI-DAPT User can check the dataset valuation and observability metrics if available.
Scenario Steps	**Configuration Time** **Step 1:** The AI-DAPT User enters the harvesting configuration. **Step 2:** The AI-DAPT User configures the data harvesting setup. The configuration includes among others the selection of data source, the provision of auth keys (if required) and accompanying information. **Step 3:** The AI-DAPT User specifies the required metadata that should be attached to the dataset. **Step 4:** The AI-DAPT Platform stores a sample dataset that will available also for the configuration of subsequent steps in the Data Pipeline or other AI-DAPT services. **Step 5:** The AI-DAPT Platform stores the harvesting setup selections in the data harvesting configuration file. **Step 6:** The AI-DAPT User views the result of the sample execution. **Step 7:** If changes are needed, the AI-DAPT User edits the configuration. Perform Steps 2-6 until satisfied with the result. **Step 8:** The AI-DAPT User finalises the configuration. **Execution Time** **Step 9:** The AI-DAPT Platform is triggered to initiate the execution of the harvesting service. **Step 10:** The AI-DAPT Platform fetches the harvesting configuration file. **Step 11:** The AI-DAPT Platform fetches data from the selected data source. **Step 12:** The AI-DAPT Platform appends the fetched data to the designated AI-DAPT dataset. **Step 13:** The AI-DAPT Platform attaches additional metadata and updates the values of metadata fields based on the latest AI-DAPT dataset version. **Step 14:** The AI-DAPT Platform stores the AI-DAPT dataset and metadata. **Step 15:** The AI-DAPT Platform updates the Dataset Catalogue. The dataset is available in the dataset catalogue for the execution of subsequent steps in the Data Pipeline or for use in other AI-DAPT Services and in AI Pipelines

	Step 16: The AI-DAPT User receives a notification about the outcome of the execution of the harvesting. **Step 17:** (optional) The AI-DAPT User can inspect the dataset valuation and observability metrics if available (SCE-04).

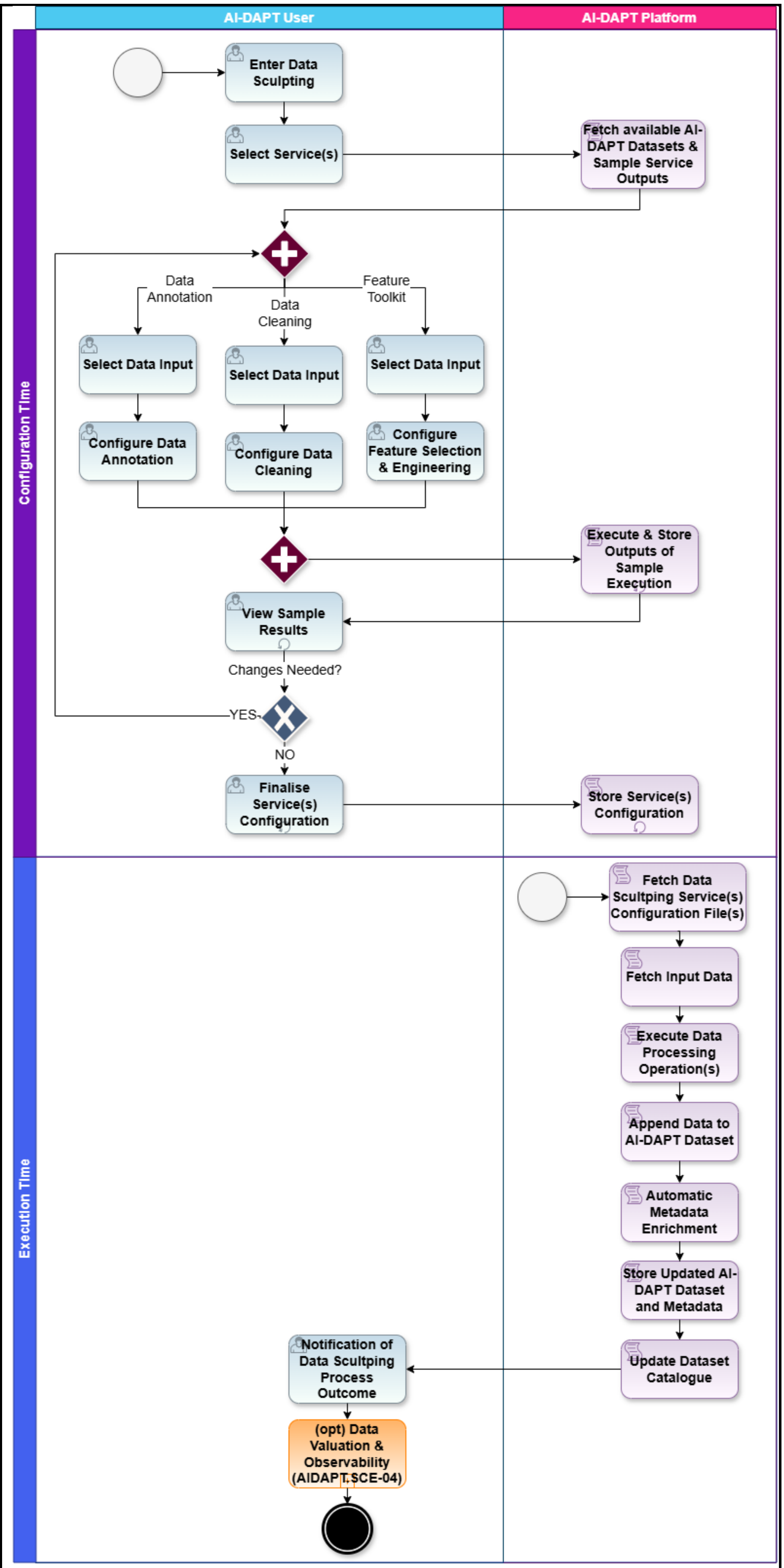
Scenario Workflow Diagram	 AI-DAPT User AI-DAPT Platform Configuration Time Execution Time Enter Harvesting Configure Harvester Service Provide Metadata Create/Update & Store Harvesting Configuration File Create & Store Sample Dataset Check Sample Changes Needed? YES NO Finalise Configuration Fetch Harvesting Configuration File Access Data Source & Fetch Data Append Data to AI-DAPT Dataset Automatic Metadata Enrichment Store Updated AI-DAPT Dataset and Metadata Update Dataset Catalogue Notification of Data Harvesting Process Outcome (opt) Data Valuation & observability (AIDAPT.SCE-04)
Users Involved	Business User, Data Scientist

Users' Benefits	The users can onboard their data to the AI-DAPT Platform, in order to utilise it in their AI operations, through a flexible retrieval mechanism supporting various retrieval methods, periodicity etc. The modular design of the harvesting service allows both the independent use the service, or its utilisation as part of a Data Pipeline.
Challenges	Obstacles in establishing connection with user data sources (e.g. no exposure to web) Specificities in the format and structure of input data specificities (e.g., proprietary or custom/non-standardised data formats)
Success Criteria	Data are successfully harvested by the AI-DAPT Platform according to user configuration. The harvested data are available for further use within the AI-DAPT Platform.

5.2.2 SCE-02: Data Sculpting

Data Sculpting - [SCE-02]	
Scenario ID	SCE-02
Scenario Name	Data Sculpting
Scenario Overview	The authenticated AI-DAPT User wants to apply the appropriate data sculpting methods over data in order to produce a dataset that is suitable for their AI operations. To do so, the AI-DAPT User enters the data sculpting services (for data annotation, cleaning, feature selection and engineering) and configures their parameters. The configuration of each service is saved in a dedicated configuration file for its execution, while a sample output is saved at configuration time and is available for preview from the AI-DAPT User. Once the configuration is finalised, the Data Sculpting services are triggered and executed according to the configuration, and their results are stored and annotated to allow the execution of other services, resulting in the final AI-DAPT dataset that is persisted in the AI-DAPT Platform, along with new metadata generated by the Documentation Engine. The AI-DAPT User receives the relevant notifications about the outcome of each Data Sculpting service execution and can inspect the dataset's valuation and observability metrics.
Scenario Steps	**<u>Configuration Time</u>** **Step 1:** The AI-DAPT User enters the Data Sculpting service(s) configuration (selection from Data Annotation Engine, Data Cleaning, Data Features Toolkit). **Step 2:** <For each service> The AI-DAPT User selects the data input of the service. It can be an AI-DAPT Dataset or the output of another service. **Step 3:** The AI-DAPT User provides the required configuration through the service's UI. **Step 4:** The AI-DAPT Platform creates and stores a sample service output. **Step 5:** The AI-DAPT Platform stores the configuration file of the specific service **Step 6:** The AI-DAPT User inspects the sample output and performs any configuration updates. **Step 7:** Repeat Steps 3-6 until satisfied with the results. **Step 8:** The AI-DAPT User finalises the configuration. **Step 9:** The AI-DAPT Platform stores the Data Sculpting service configuration **<u>Execution Time</u>** **Step 10:** The AI-DAPT Platform is triggered to initiate the execution of the Data Sculpting service(s). **Step 11:** The AI-DAPT Platform fetches the data sculpting service configuration file. **Step 12:** The AI-DAPT Platform fetches the selected input data. **Step 13:** The AI-DAPT Platform executes the data sculpting process according to the configuration.

	Step 14: The Ai-DAPT Platform loads the data to the designated AI-DAPT dataset. **Step 15:** The AI-DAPT Platform attaches additional metadata and updates the values of metadata fields based on the latest AI-DAPT dataset version. **Step 16:** The AI-DAPT Platform stores the AI-DAPT dataset and metadata. **Step 17:** The AI-DAPT Platform updates the Dataset Catalogue. The dataset is available in the dataset catalogue for the execution of subsequent steps in the Data Pipeline or for use in other AI-DAPT Services and in AI Pipelines **Step 18:** The AI-DAPT User receives a notification about the outcome of the execution of the data sculpting service(s). **Step 19:** (optional) The AI-DAPT User can inspect the dataset valuation and observability metrics if available (SCE-04).

Scenario Workflow Diagram	

AI-DAPT User
AI-DAPT Platform
Configuration Time
Execution Time
Enter Data Sculpting
Select Service(s)
Fetch available AI-DAPT Datasets & Sample Service Outputs
Data Annotation
Data Cleaning
Feature Toolkit
Select Data Input
Select Data Input
Select Data Input
Configure Data Annotation
Configure Data Cleaning
Configure Feature Selection & Engineering
Execute & Store Outputs of Sample Execution
View Sample Results
Changes Needed?
YES
NO
Finalise Service(s) Configuration
Store Service(s) Configuration
Fetch Data Scultping Service(s) Configuration File(s)
Fetch Input Data
Execute Data Processing Operation(s)
Append Data to AI-DAPT Dataset
Automatic Metadata Enrichment
Store Updated AI-DAPT Dataset and Metadata
Update Dataset Catalogue
Notification of Data Scultping Process Outcome
(opt) Data Valuation & Observability (AIDAPT.SCE-04)

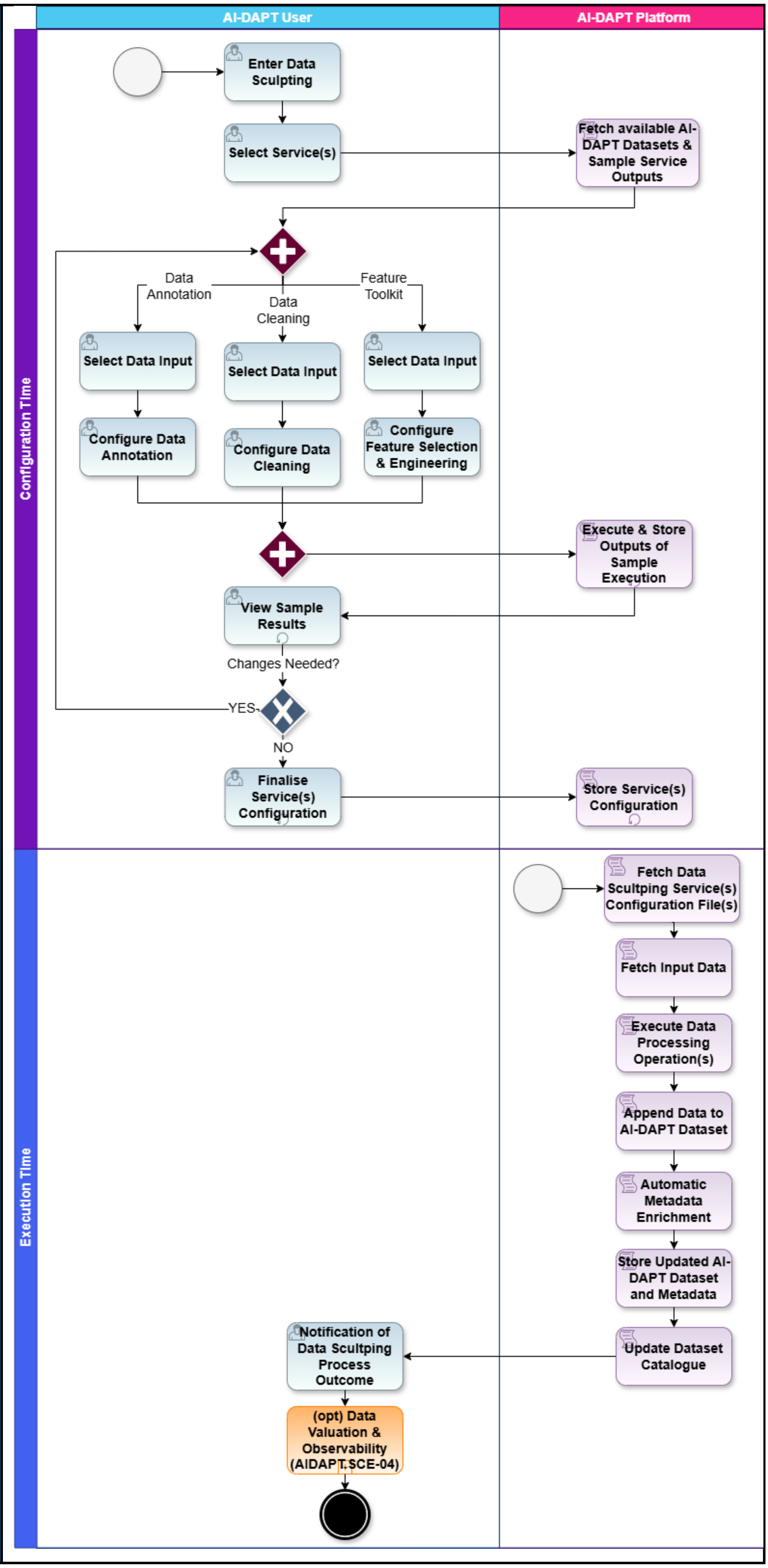
Users Involved	Business User (possibly limited expertise for some services/advanced configurations), Data Scientist
Users' Benefits	The users can apply data sculpting processes (including data cleaning operations, semantic reconciliation and feature engineering) over their data in order to optimise them for their AI operations in AI-DAPT. The modular design of the data sculpting services allows both the independent use the services, or their utilisation as part of a Data Pipeline.
Challenges	Case-agnostic/generic data sculpting services might not be able to address custom/proprietary/non-standardised data with format and structure specificities. Provision of standards-based data models addressing effectively any data harmonisation needs. Effectively introduce automation to the data sculpting services.
Success Criteria	Data are successfully processed by the AI-DAPT Platform according to user configuration. The processed data are available for further use within the AI-DAPT Platform.

5.2.3 SCE-03: Synthetic Data Generation

Synthetic Data Generation - [SCE-03]	
Scenario ID	SCE-03
Scenario Name	Synthetic Data Generation
Scenario Overview	The authenticated AI-DAPT User wants to create synthetic data, either to utilise them in their AI Pipelines or for other purposes in their applications. For this purpose, the AI-DAPT User creates a Synthetic Data Generation project. The AI-DAPT User can explore their original data (if they are going to use existing data for synthetic data generation) to familiarise with the dataset's characteristics. Afterwards they can proceed to the selection of the synthetic data generation process and methods and the configuration of the required parameters. Once ready with configuration, the synthetic data generation project is executed. The AI-DAPT User can view the characteristics of the produced dataset, compare it with the original (if applicable) in order to verify the suitability of the data for their purposes, or to perform any required adjustments in the generation configuration and re-execute the project, until they get a synthetic dataset that fits their needs. The AI-DAPT User can export the generated dataset, while it is also registered in the AI-DAPT catalogue for further use in other AI Pipelines or consumption by the external applications. The created synthetic data generation project is available for future re-use.
Scenario Steps	**Configuration Time** **Step 1:** The AI-DAPT User enters the Synthetic Data Generation Engine UI and creates a new Synthetic Data Generation Project selecting the main generation path (i.e. option A: from scratch or, option B: based on existing data) or re-uses an existing Synthetic Data Generation Project (directly edit, or copy configuration and accompanying information to new project). **Step 2:** If the AI-DAPT User has selected to re-use an existing Synthetic Data Generation Project, all previous configurations are pre-loaded. **Step 3:** If the AI-DAPT User has selected to use existing data, she should indicate an existing AI-DAPT dataset from the catalogue to be used. **Step 4:** The AI-DAPT User configures/edits the synthetic data generation properties. The available options depend on the selected path (i.e. from scratch or from existing data), and may include among others: the selection of mode, generation technique and hyper-parameters, as well as the synthetic dataset properties (e.g. number of rows). **Step 5:** The AI-DAPT User completes the configuration and triggers the execution of the Synthetic Data Generation Project **Step 6:** The AI-DAPT Platform stores the intermediary Synthetic Data Generation Project, configuration and sample data (if uploaded), along with the appropriate metadata so that they are available for future re-use.

	Execution Time **Step 7:** The AI-DAPT Platform is triggered to execute the Synthetic Data Generation Project **Step 8:** The AI-DAPT Platform executes the Synthetic Data Generation Project based on the provided configuration and a new synthetic dataset is generated. **Step 9:** The AI-DAPT Platform stores the intermediary synthetic dataset along with the generation configuration file. **Step 10:** The AI-DAPT User receives notification about the results of the execution of the Synthetic Data Generation Project **Step 11:** The AI-DAPT User verifies the utility of the generated dataset through data preview and basic visualisations. **Step 12:** If the AI-DAPT User is satisfied with the results she finalises the configuration and proceeds to the following steps, otherwise repeat Steps 4-11. **Step 13:** The AI-DAPT Platform stores the latest Synthetic Data Generation Project, configuration and sample data (if uploaded), along with the appropriate metadata so that they are available for future re-use. **Step 14:** The AI-DAPT Platform stores the AI-DAPT dataset and metadata.

Scenario Workflow Diagram


AI-DAPT User
AI-DAPT Platform
Configuration Time
Enter Data Sculpting
Select Service(s)
Fetch available AI-DAPT Datasets & Sample Service Outputs
Data Annotation
Data Cleaning
Feature Toolkit
Select Data Input
Select Data Input
Select Data Input
Configure Data Annotation
Configure Data Cleaning
Configure Feature Selection & Engineering
Execute & Store Outputs of Sample Execution
View Sample Results
Changes Needed?
YES
NO
Finalise Service(s) Configuration
Store Service(s) Configuration
Execution Time
Fetch Data Sculptping Service(s) Configuration File(s)
Fetch Input Data
Execute Data Processing Operation(s)
Append Data to AI-DAPT Dataset
Automatic Metadata Enrichment
Store Updated AI-DAPT Dataset and Metadata
Update Dataset Catalogue
Notification of Data Sculptping Process Outcome
(opt) Data Valuation & Observability (AIDAPT.SCE-04)

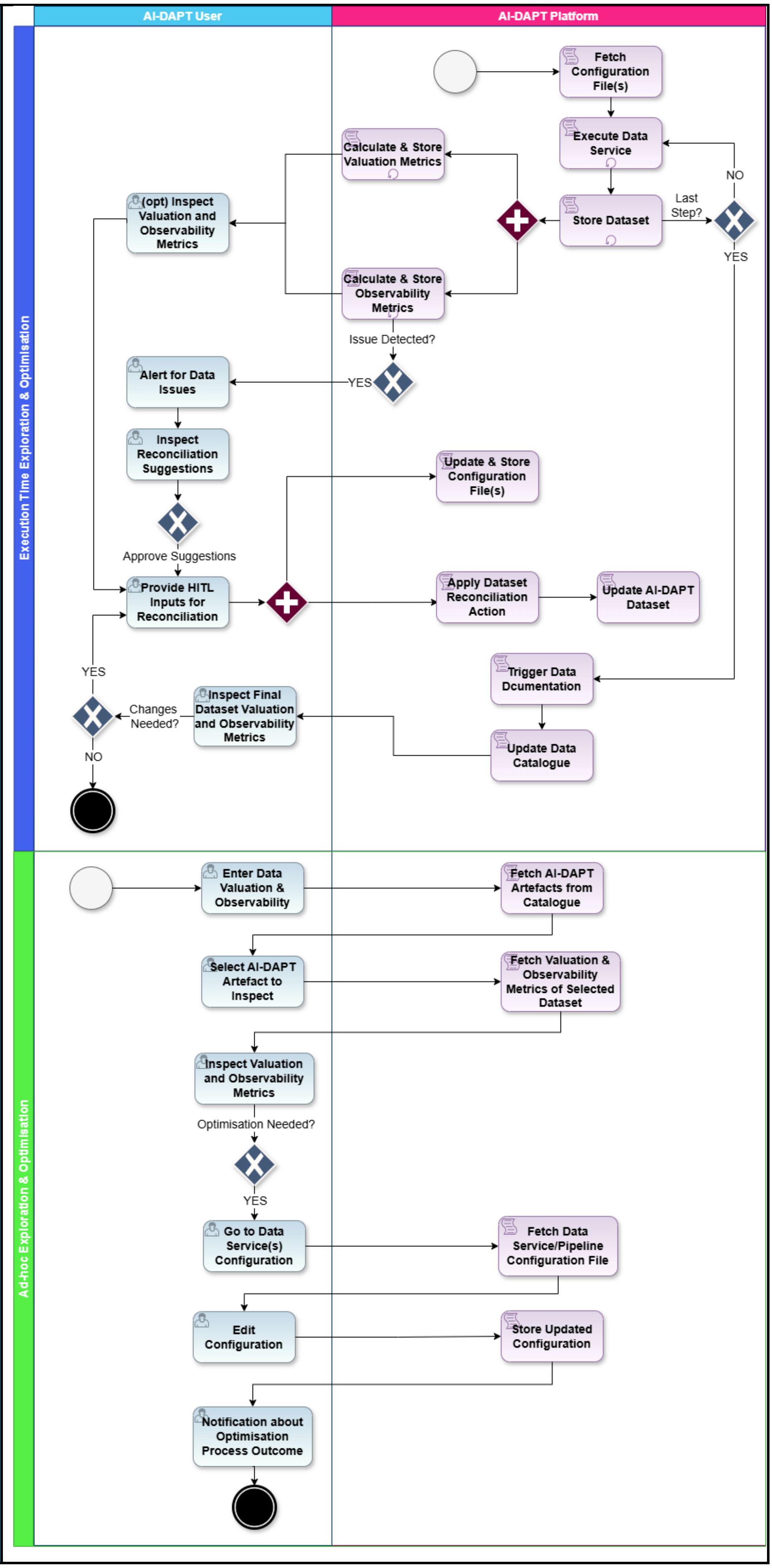
Users Involved	Business User (possibly limited expertise for advanced configurations), Data Scientist
Users' Benefits	The users can create synthetic data in a flexible manner and meeting their data standards, in order to augment their AI operations. Synthetic data are useful in cases where original data are not of sufficient volume, diversity or otherwise inappropriate for further use (e.g. constrains due to sensitive and/or private nature). The modular design of the synthetic data generation service allows both the independent use the service, or its utilisation as part of a Data Pipeline.
Challenges	Definition of generic/use-case-agnostic synthetic data utility assessment framework might be challenging.
Success Criteria	Synthetic data are successfully generated by the AI-DAPT Platform according to user configuration. The synthetic data are available for further use within the AI-DAPT Platform.

5.2.4 SCE-04: Data Valuation, Observability & Optimisation

Data Valuation, Observability & Optimisation - [SCE-04]	
Scenario ID	SCE-04
Scenario Name	Data Valuation, Observability & Optimisation
Scenario Overview	The authenticated AI-DAPT User wants to ensure the high quality of their AI-DAPT datasets and perform the required reconciliation actions to improve weak points. For this purpose, the AI-DAPT User can check ad-hoc the data valuation metrics extracted from a specific dataset and view comparative analysis against selected ground-truth data. The AI-DAPT User can also monitor the health status of the dataset throughout its entire lifecycle through observability metrics and logs, while they are alerted whenever an issue is detected (e.g. data drift, anomalies etc.). After inspecting the available information, if the problem lies on the side of the AI-DAPT data handling, the AI-DAPT User can revise the configuration of their Data Pipeline or of a specific data manipulation service to address data quality or data performance issues.
Scenario Steps	**Execution Time Exploration & Optimisation** **Step 1:** The AI-DAPT Platform is triggered to initiate the execution of data services. **Step 2:** The AI-DAPT Platform retrieves the configuration files of the data services. **Step 3:** <For each service x> The AI-DAPT Platform executes the Service x according to configuration and utilising the designated data inputs. **Step 4:** The AI-DAPT Platform stores the result of the execution of the data manipulation service <x> in the designated AI-DAPT dataset **Step 5:** The AI-DAPT Platform calculates the valuation, monitoring and observability metrics. **Step 6:** If any issue of data degradation is identified, the AI-DAPT Platform generates an informative alert. **Step 7:** The AI-DAPT User reviews the alert and relevant logs and approves/rejects the suggested reconciliation actions. (e.g. omit specific dataset update and reconfigure pipeline or apply data correction actions). **Step 8:** The AI-DAPT Platform receives the user response and updates the dataset or the pipeline configuration accordingly. **Step 9:** Repeat Steps 3 – 8 until all data services are executed **Step 10:** The AI-DAPT Platform attaches additional metadata and updates the values of metadata fields based on the latest AI-DAPT dataset version **Step 11:** The AI-DAPT Platform stores the AI-DAPT dataset and metadata.

	Ad-hoc Exploration & Optimisation **Step 12:** The AI-DAPT User views the valuation and observability metrics of a specific AI-DAPT artefact (e.g. dataset) **Step 13:** The AI-DAPT Platform fetches the artefact's metrics. **Step 14:** The AI-DAPT User inspects the valuation and observability metrics **Step 15:** The AI-DAPT User makes the appropriate data service reconfigurations if needed. **Step 16:** The AI-DAPT User finalises the Data Pipeline configuration **Step 17:** The AI-DAPT Platform stores the updated data service configuration files. **Step 18:** The AI-DAPT User receives a notification about the outcome of the optimisation process.

Scenario Workflow Diagram

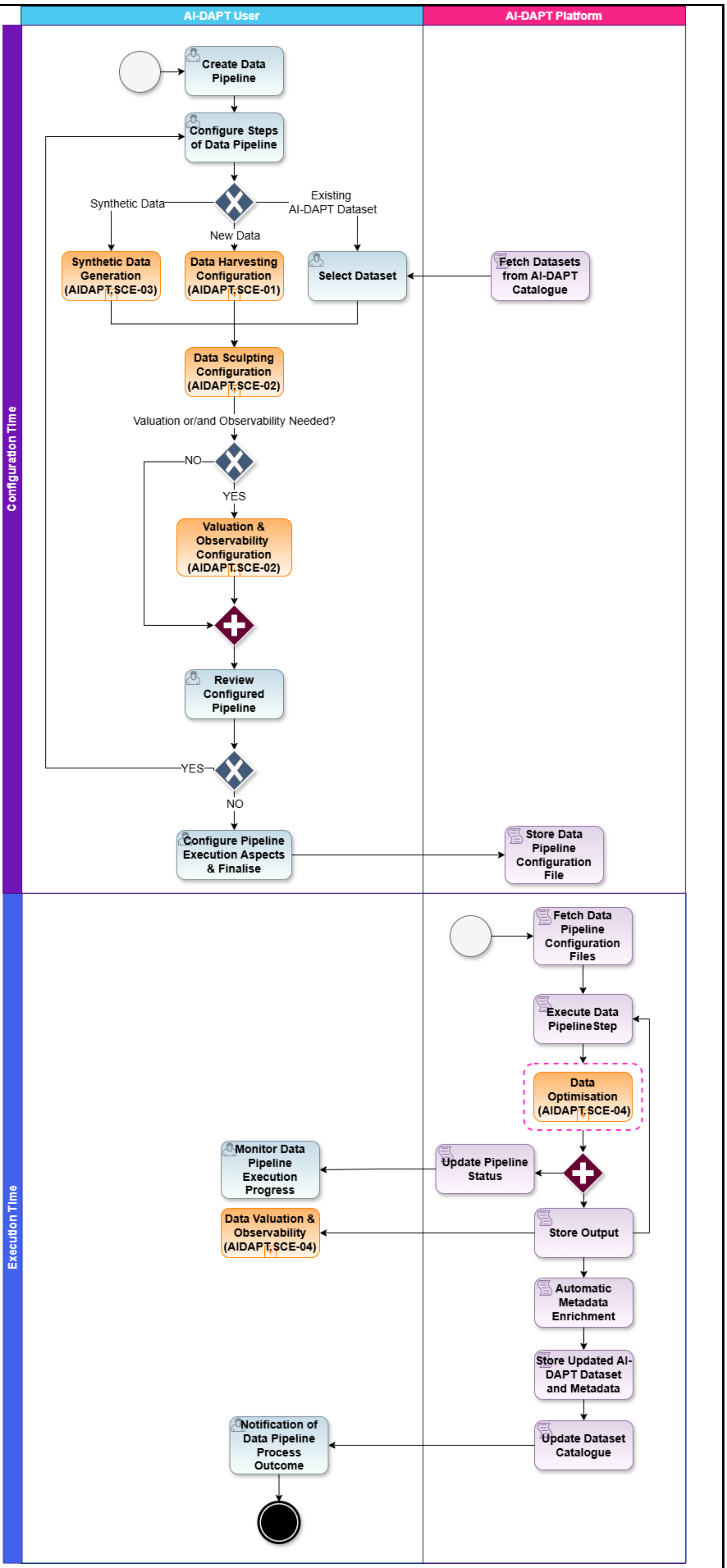

AI-DAPT User
AI-DAPT Platform
Execution Time Exploration & Optimisation
Fetch Configuration File(s)
Execute Data Service
Calculate & Store Valuation Metrics
NO
Last Step?
Store Dataset
YES
(opt) Inspect Valuation and Observability Metrics
Calculate & Store Observability Metrics
Issue Detected?
Alert for Data Issues
YES
Inspect Reconciliation Suggestions
Update & Store Configuration File(s)
Approve Suggestions
Provide HITL Inputs for Reconciliation
Apply Dataset Reconciliation Action
Update AI-DAPT Dataset
Trigger Data Dcumentation
YES
Changes Needed?
Inspect Final Dataset Valuation and Observability Metrics
NO
Update Data Catalogue
Ad-hoc Exploration & Optimisation
Enter Data Valuation & Observability
Fetch AI-DAPT Artefacts from Catalogue
Select AI-DAPT Artefact to Inspect
Fetch Valuation & Observability Metrics of Selected Dataset
Inspect Valuation and Observability Metrics
Optimisation Needed?
YES
Go to Data Service(s) Configuration
Fetch Data Service/Pipeline Configuration File
Edit Configuration
Store Updated Configuration
Notification about Optimisation Process Outcome

Users Involved	Business User, Data Scientist
Users' Benefits	The users monitor the high quality of their data throughout their lifecycle and can optimise the data services configuration to address data-quality issues. The modular design of the data valuation and observability services allows both the independent use of the services, or their utilisation as part of a Data Pipeline.
Challenges	Definition of generic/use-case-agnostic valuation assessment framework might be challenging. Provision of optimisation suggestions might be challenging.
Success Criteria	The calculated valuation and observability metrics for an AI-DAPT Dataset are available to the AI-DAPT User. Alerts are generated according to user preferences. Data optimisation suggestions are provided by the AI-DAPT Platform. Data Pipelines/Services are adapted and re-executed according to user optimisation configuration.

5.2.5 SCE-05: Data Pipeline Design & Execution

Data Pipeline Design & Execution - [SCE-05]	
Scenario ID	SCE-05
Scenario Name	Data Pipeline Design & Execution
Scenario Overview	The authenticated AI-DAPT User wants to create a Data Pipeline that will be executed based on their specific needs. For this purpose, the AI-DAPT User creates a custom Data Pipeline, and adds to the Pipeline the data pre-processing steps that should be performed (selection from Data Harvesting, Data Sculpting, Synthetic Data Generation). The AI-DAPT User can also infuse data valuation, observability and optimisation steps in the Pipeline to further enhance it. The AI-DAPT User is facilitated to take informed configuration decisions through the experimentation services of AI-DAPT. Once ready, the AI-DAPT User finalises the pipeline configuration. The Data Pipeline is executed according to the configuration, with intermediate outputs stored and annotated to allow the execution of the next step of the pipeline, resulting in the final AI-DAPT dataset that is persisted in the AI-DAPT Platform, along with new metadata. If optimisation has been infused in the pipeline, the appropriate adaptations are made to the pipeline at production time. The AI-DAPT User has an overview of the Data Pipeline Execution status per step and receives the relevant notification upon finalisation of the execution.
Scenario Steps	**Configuration Time** **Step 1:** The AI-DAPT User creates a new Data Pipeline. **Step 2:** The AI-DAPT User adds and configures a data harvesting step (SCE-01) or selects an already available dataset from the AI-DAPT Catalogue. **Step 3:** The AI-DAPT User adds and configures any data sculpting service(s) that should be applied on the data (selection from Data Annotation Engine, Data Cleaning, Data Features Toolkit) (SCE-02). **Step 4:** If needed, the AI-DAPT User adds synthetic data generation step (if needed) (SCE-03). **Step 5:** If changes are needed the AI-DAPT User makes the appropriate reconfigurations. Repeat Steps 2 – 4 until satisfied with the results. **Step 6:** The AI-DAPT User adds data valuation and observability steps to the Data Pipeline. **Step 7:** The AI-DAPT User configures the dataset metadata. **Step 8:** The AI-DAPT User configures the Data Pipeline execution aspects and finalises the Pipeline. **Step 9:** The AI-DAPT Platform stores the Data Pipeline configuration.

	Execution Time **Step 10:** The AI-DAPT Platform is triggered to initiate the execution of the Data Pipeline according to the Data Pipeline execution configuration. **Step 11:** The AI-DAPT Platform retrieves the configuration files of the Data Pipeline and the individual services. **Step 12:** The AI-DAPT Platform executes each step of the Data Pipeline. **Step 13:** The AI-DAPT Platform stores the output of the execution of each step. **Step 14:** The AI-DAPT Platform shows the progress of the execution of each step and the results of the intermediate steps feed the valuation & observability services. **Step 15:** Repeat Steps 12-14 until all services of the Pipeline are executed **Step 16:** The AI-DAPT Platform attaches additional metadata and updates the values of metadata fields based on the latest AI-DAPT dataset version. **Step 17:** The AI-DAPT Platform stores the AI-DAPT dataset and metadata. **Step 18:** The dataset is available for use in other AI-DAPT Services and in AI Pipelines. **Step 19:** The AI-DAPT User monitors and receives notifications about the results of the execution of the Data Pipeline.

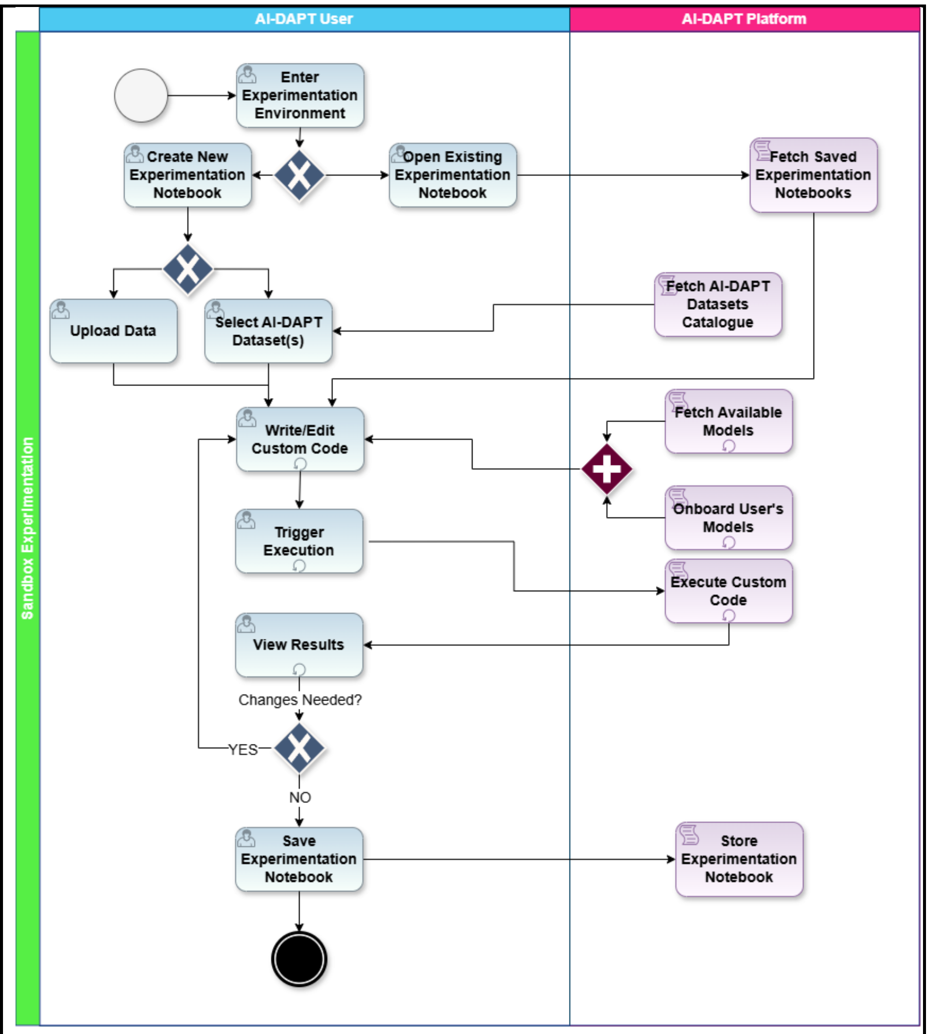
Scenario Workflow Diagram	

AI-DAPT User
AI-DAPT Platform
Configuration Time
Execution Time
Create Data Pipeline
Configure Steps of Data Pipeline
Synthetic Data
Existing AI-DAPT Dataset
New Data
Synthetic Data Generation (AIDAPT.SCE-03)
Data Harvesting Configuration (AIDAPT.SCE-01)
Select Dataset
Fetch Datasets from AI-DAPT Catalogue
Data Sculpting Configuration (AIDAPT.SCE-02)
Valuation or/and Observability Needed?
NO
YES
Valuation & Observability Configuration (AIDAPT.SCE-02)
Review Configured Pipeline
YES
NO
Configure Pipeline Execution Aspects & Finalise
Store Data Pipeline Configuration File
Fetch Data Pipeline Configuration Files
Execute Data Pipeline Step
Data Optimisation (AIDAPT.SCE-04)
Monitor Data Pipeline Execution Progress
Update Pipeline Status
Data Valuation & Observability (AIDAPT.SCE-04)
Store Output
Automatic Metadata Enrichment
Store Updated AI-DAPT Dataset and Metadata
Notification of Data Pipeline Process Outcome
Update Dataset Catalogue

Users Involved	Business User (possibly limited expertise for some services/advanced configurations), Data Scientist
Users' Benefits	Users can create custom Data Pipelines utilising in a flexible manner the AI-DAPT Data services (harvesting, synthetic data generation, data sculpting). The Users stay in control of their pipelines at execution time and ensure high quality of data, through alerts, observability and valuation metrics, allowing them to monitor closely both the data quality and the pipeline status per se.
Challenges	Interplay of the self-standing data manipulation and monitoring/optimisation services with Pipeline Designer and Execution Engine. Translation of complex data manipulation services in Data Pipeline blocks/operators.
Success Criteria	Data are successfully harvested, processed and monitored by the AI-DAPT Platform according to the defined Data Pipeline. The data outputs of a Pipeline are available for further use within the AI-DAPT Platform.

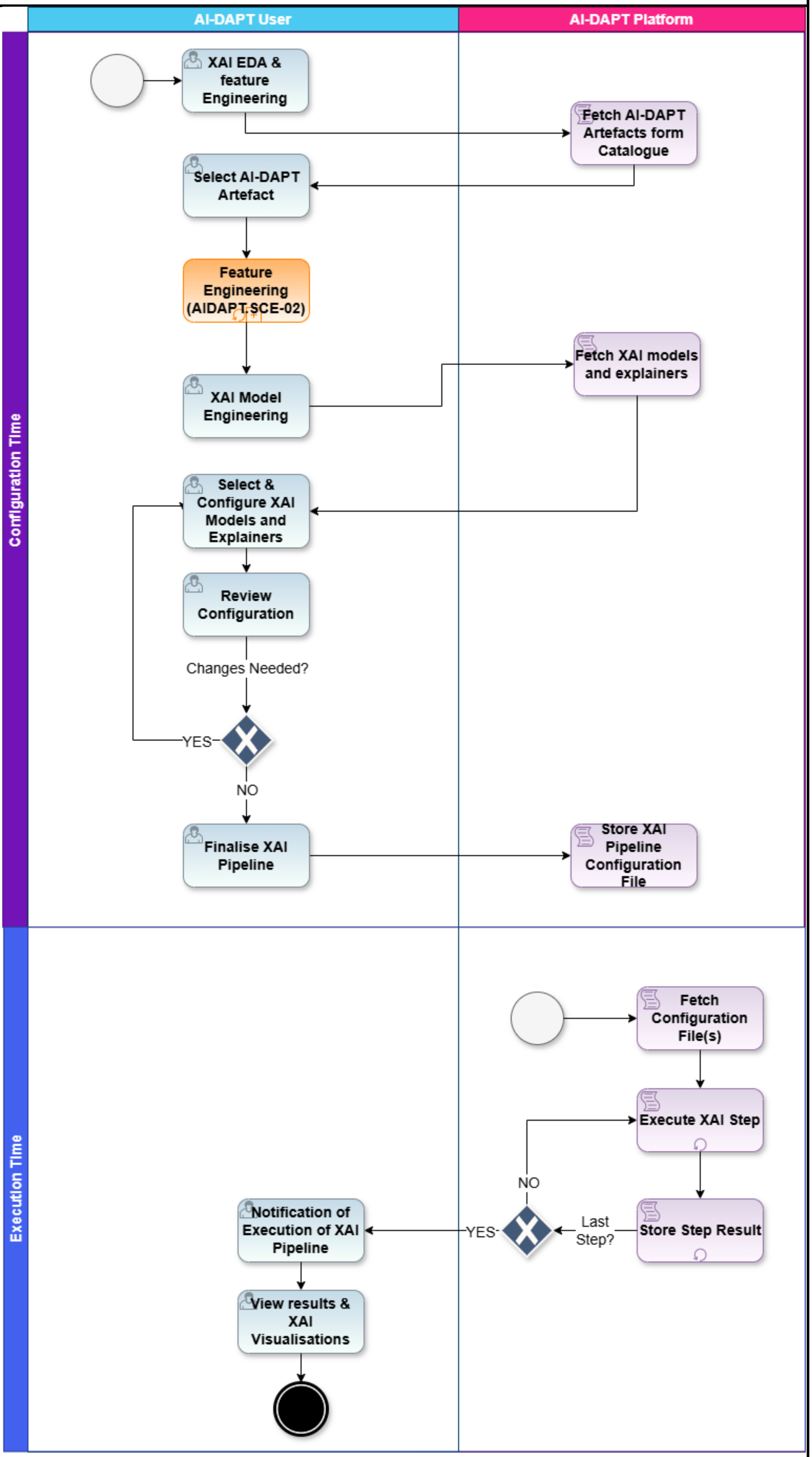
5.2.6 SCE-06: Data & Model Interactive Experimentation

Data & Model Interactive Experimentation - [SCE-06]	
Scenario ID	SCE-06
Scenario Name	Data & Model Interactive Experimentation
Scenario Overview	An AI-DAPT User with the relevant development/data science background wants to experiment with their own models and data pre-processing code before they create their production pipelines. They go to the experimentation environment where they can write their own code, execute it and see the results. They can store their experimentation notebook for future reference and use.
Scenario Steps	**Sandbox Experimentation** **Step 1:** The AI-DAPT User enters the experimentation environment. **Step 2:** The AI-DAPT User selects either to create a new experimentation notebook or continue an existing one. **Step 3:** The AI-DAPT User selects the AI-DAPT Dataset(s) that should be used or/and uploads additional data. **Step 4:** The AI-DAPT User experiments with custom code and inspects the results. **Step 5:** Repeat Step 4 until satisfied with the results. **Step 6:** The AI-DAPT User saves the experimentation notebook. **Step 7:** The AI-DAPT Platform stored the notebook for future reuse.

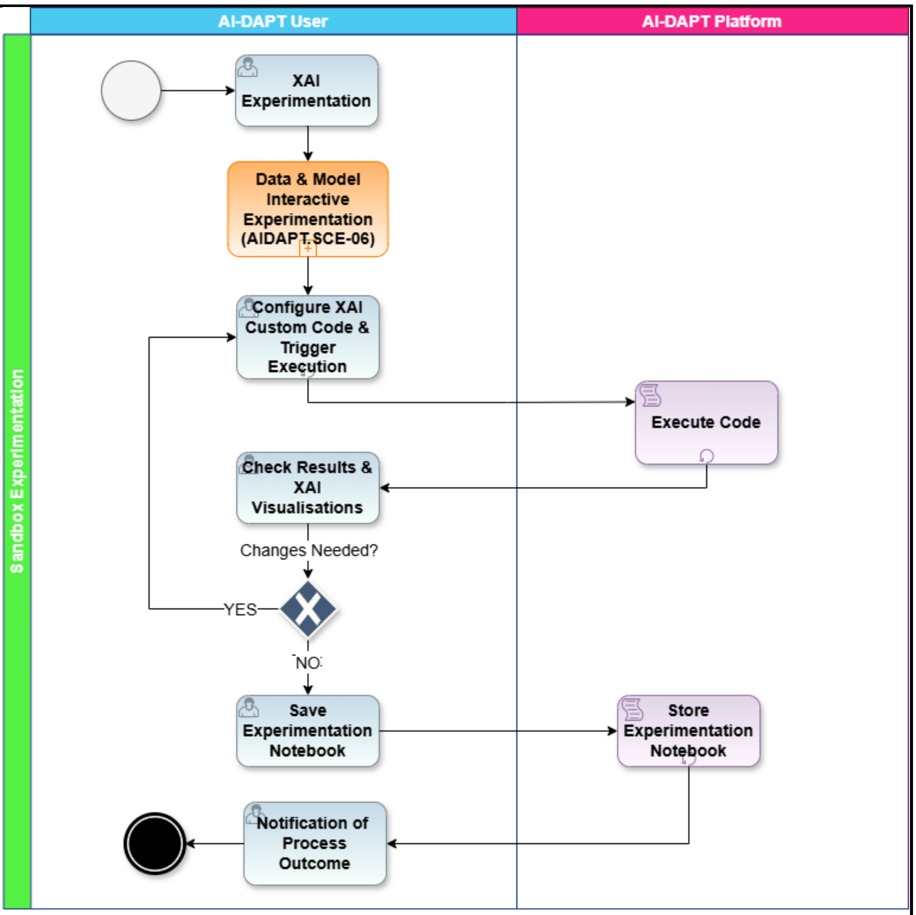
Scenario Workflow Diagram	 AI-DAPT User AI-DAPT Platform Sandbox Experimentation Enter Experimentation Environment Create New Experimentation Notebook Open Existing Experimentation Notebook Fetch Saved Experimentation Notebooks Fetch AI-DAPT Datasets Catalogue Upload Data Select AI-DAPT Dataset(s) Write/Edit Custom Code Fetch Available Models Onboard User's Models Trigger Execution Execute Custom Code View Results Changes Needed? YES NO Save Experimentation Notebook Store Experimentation Notebook
Users Involved	Business User (possibly limited expertise for sandbox experimentation), Data Scientist
Users' Benefits	The users can go beyond the models and data processing methods available through AI-DAPT and experiment with custom models and code in a sandboxed environment.
Challenges	Interplay of Experimentation Engine within the AI-DAPT Framework. Cybersecurity issues (e.g., malicious custom code).
Success Criteria	The AI-DAPT User can upload/write custom code, execute it and see the results in the AI-DAPT Platform. The AI-DAPT User can find and edit previous experimentation notebooks.

5.2.7 SCE-07: Model Explanation

Model Explanation - [SCE-07]	
Scenario ID	SCE-07
Scenario Name	Model Explanation
Scenario Overview	The authenticated AI-DAPT User wants to take advantage of the available XAI models and techniques. At design time the AI-DAPT User explores the dataset that will be used in their AI operations. Through interactive experimentation and feature engineering, the AI-DAPT User gains insights over the input data and their characteristics and identifies the most appropriate features for the derivation of explanations. The AI-DAPT User afterwards builds an AI Pipeline according to their needs, selecting from the available models the ones that are applicable to their use case. Some of the available models can either be inherently interpretable (i.e., white-box models), while the AI-DAPT User can also add explainer blocks to the XAI Pipeline (e.g. supporting SHAP, LIME). Once satisfied with the designed AI Pipeline, the AI-DAPT User finalises its configuration and it goes to production, where the AI-DAPT User can view the outcomes and explanations in intuitive visualisations to gain insights in the model outcomes.
Scenario Steps	**XAI Techniques in AI** **Step 1:** The AI-DAPT User utilises the appropriate data pre-processing services (e.g. Feature Engineering in SCE-2) to explore their data and select the most suitable features. They can also perform Exploratory Data Analytics (SCE-6). **Step 2:** The AI-DAPT User selects white-box or black-box models. **Step 3:** In case of black-box models, the AI-DAPT User selects one of the available explainability methods. **Step 4:** Repeat steps 2-4 until satisfied with the results. **Step 5:** The AI-DAPT User finalises the configuration. **Step 6:** The AI-DAPT Platform stores the finalised service / pipeline configuration file(s). **Step 7:** The AI-DAPT Platform is triggered to initiate the execution of the AI Pipeline according to the Pipeline execution configuration. **Step 8:** The AI-DAPT Platform executes the model steps according to the configuration. **Step 9:** The AI-DAPT Platform executes the explainer steps and stores the results. **Step 10:** The AI-DAPT User is notified about the completion of the execution. **Step 11:** The AI-DAPT User inspects the training and evaluation results and explanations. **XAI Sandbox Experimentation** **Step 12:** If experimentation is to be performed outside a Pipeline, the User can perform the design steps of SCE-06, but this time they focus on the XAI models and explainers that will have as an output the relevant explanations.

	Step 13: The AI-DAPT User inspects the results and sample explanations **Step 14:** Repeat until satisfied with the results. **Step 15:** The AI-DAPT User stores the experimentation notebook for future use.
Scenario Workflow Diagram	

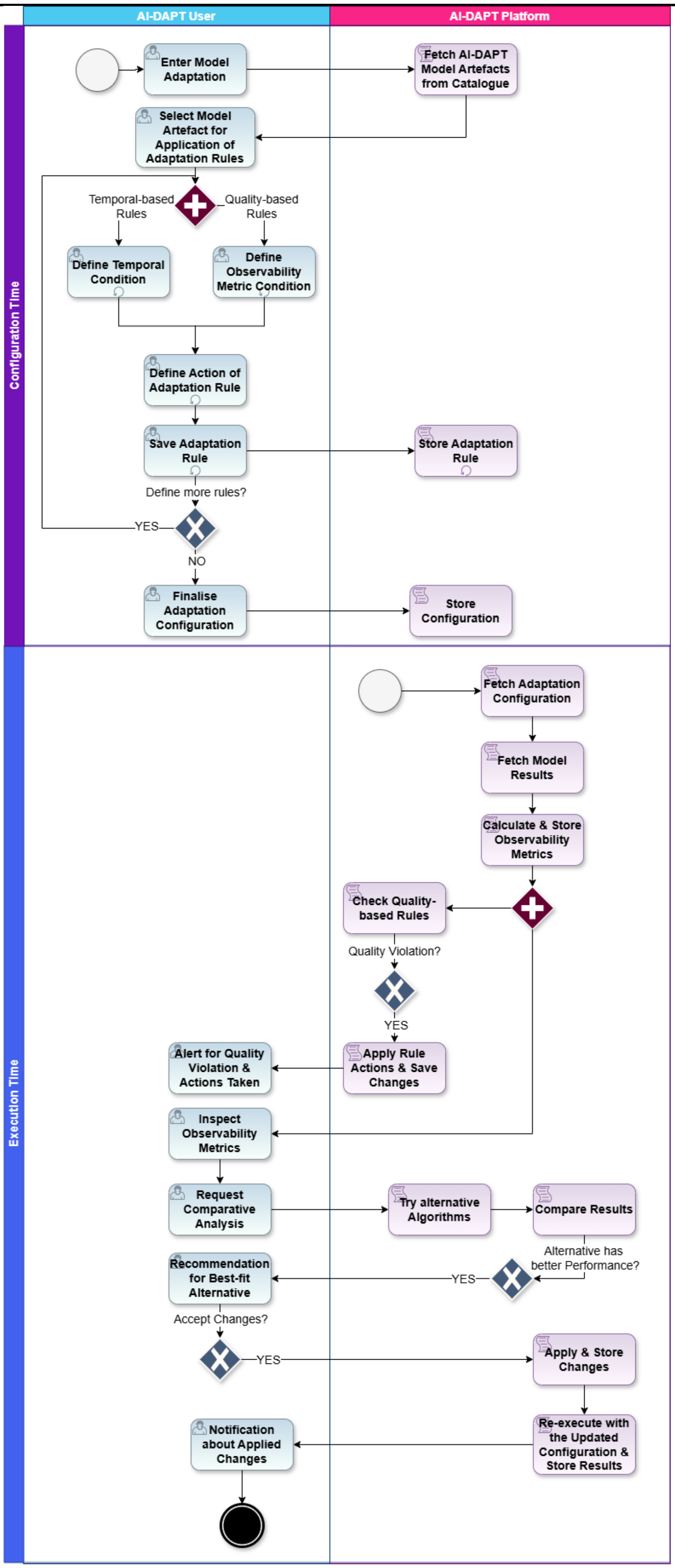
AI-DAPT User
AI-DAPT Platform
Configuration Time
XAI EDA & feature Engineering
Fetch AI-DAPT Artefacts form Catalogue
Select AI-DAPT Artefact
Feature Engineering (AIDAPT SCE-02)
Fetch XAI models and explainers
XAI Model Engineering
Select & Configure XAI Models and Explainers
Review Configuration
Changes Needed?
YES
NO
Finalise XAI Pipeline
Store XAI Pipeline Configuration File
Execution Time
Fetch Configuration File(s)
Execute XAI Step
NO
Notification of Execution of XAI Pipeline
YES
Last Step?
Store Step Result
View results & XAI Visualisations

	 AI-DAPT User AI-DAPT Platform Sandbox Experimentation XAI Experimentation Data & Model Interactive Experimentation (AIDAPT.SCE-06) Configure XAI Custom Code & Trigger Execution Execute Code Check Results & XAI Visualisations Changes Needed? YES NO Save Experimentation Notebook Store Experimentation Notebook Notification of Process Outcome
Users Involved	Data Scientist
Users' Benefits	Users can utilise the available XAI techniques (i.e. interpretable models, visualisations, explainers) in order to understand the outcomes of their models and AI Pipelines. The AI-DAPT approach to XAI is multifaceted, spanning from the exploratory analysis of data (though the feature engineering and experimentation engines), infusion of AI Pipelines with XAI through the appropriate models and explainers in the AI-DAPT library, the provision of XAI visualisation, but also allowing further experimentation for data scientists to use their custom XAI code.
Challenges	Interplay of XAI, Experimentation and EDA services. Design of case-agnostic/generic XAI models. Cybersecurity issues (e.g., malicious custom code).
Success Criteria	XAI models and explainers are available in the AI-DAPT knowledge base to be used in AI Pipelines. The AI-DAPT User can view the results of XAI methods in visualisations.

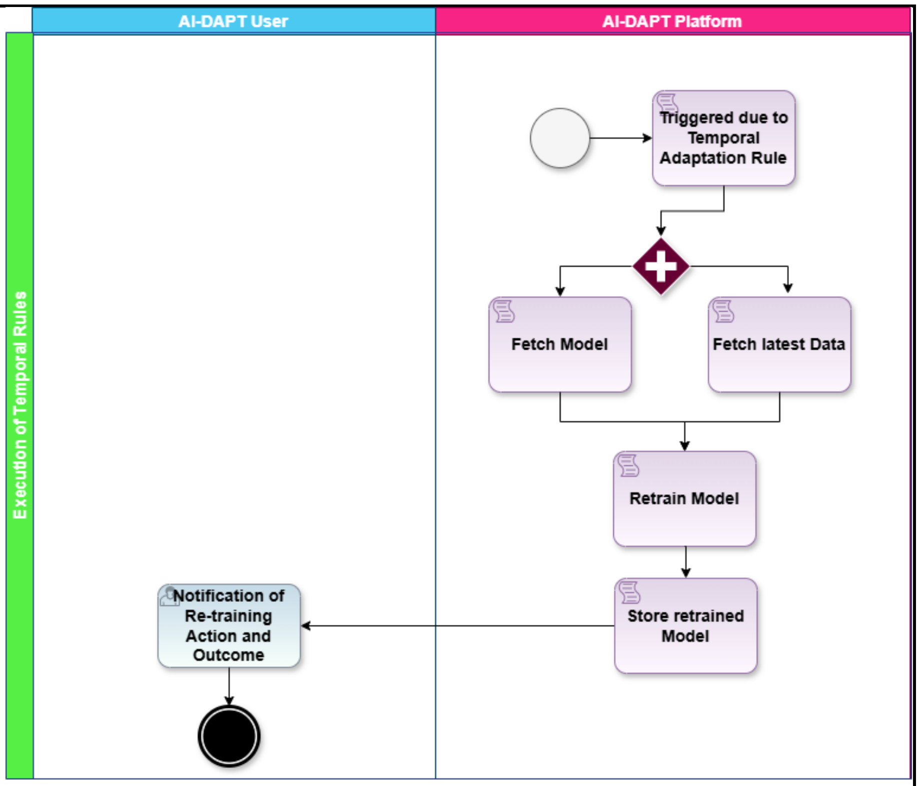
5.2.8 SCE-08: Model Observation & Adaptation

Model Observation & Adaptation - [SCE-08]	
Scenario ID	SCE-08
Scenario Name	Model Observation & Adaptation
Scenario Overview	The authenticated AI-DAPT User wants to ensure the high quality of the results of their AI Pipelines through model monitoring and the application of adaptive learning techniques. For this purpose, during the AI Pipeline configuration, the AI-DAPT User can specify rules for model adaptation based on specific model observability thresholds. The AI-DAPT User can define the actions that should be taken when model degradation is detected (e.g. replacement of model in pipeline with alternative model), or even based on fixed periodicity (e.g. retrain every x months), while the AI-DAPT User will receive alerts if specific quality aspects are not satisfied during the production phase of the Pipeline or any adaptation actions are applied. During the production phase, the outputs of the AI Pipelines are continuously monitored. The observability metrics are available for the AI-DAPT user to inspect in an ad-hoc manner, while the appropriate informative alerts are generated whenever model degradation issues are detected. The AI-DAPT Platform applies the adaptation actions defined by the AI-DAPT User. Furthermore, the AI-DAPT Platform can assist in model improvement by providing recommendations to the AI-DAPT User for adaptive learning (e.g. recommendation of a best-fit alternative model). The AI-DAPT User inspects the recommendation and approves or rejects it. Afterwards the AI-DAPT Platform applies any accepted changes.
Scenario Steps	**Configuration Time** **Step 1:** The AI-DAPT User is authenticated by the AI-DAPT Platform. **Step 2:** The AI-DAPT User selects the model where the adaptation rules should be applied. **Step 3:** The AI-DAPT User creates model adaptation rules based on observability metrics thresholds. **Step 4:** The AI-DAPT User defines the adaptation rules: 1. Actions that should be performed on the AI Pipeline for model adaptation if the defined observability thresholds of good quality are violated. 2. The AI-DAPT User creates periodical adaptive learning rules, that are applied based on temporal aspects. **Step 5:** The AI-DAPT user finalises the adaptation configuration. **Step 6:** The AI-DAPT Platform stores the adaptation configuration. **Execution Time** **Step 7:** The AI-DAPT Platform is triggered to execute the AI Pipeline according to pipeline configuration.

	Step 8: The AI-DAPT Platform calculates the observability metrics of the output model. **Step 9:** The AI-DAPT Platform checks the defined adaptation rules. If any of the rules should be applied based on the latest observability metrics, then the AI-DAPT Platform performs the relevant update in the pipeline and it is re-executed. **Step 10:** The AI-DAPT User receives an alert about the rule application and applied changes. **Step 11:** The AI-DAPT Platform performs experimentation with alternative models upon User request and if it finds a model that is performing better that the one currently selected, it generates an informative alert for the user. **Step 12:** The AI-DAPT User inspects the recommendation and accepts or rejects it. **Step 13:** If the AI-DAPT User accepts the recommendation, the AI-DAPT Platform applies the changes in the AI Pipeline and re-executes it. **Step 14:** The output of the pipeline is stored. **Step 15:** The AI-DAPT User receives a notification about the outcome of the execution. **Temporal-based Rule** **Step 16:** The AI-DAPT Platform is triggered to perform retraining due to a temporal rule. **Step 17:** The AI-DAPT Platform fetches the latest data. **Step 18:** The AI-DAPT Platform retrains the model with new data. **Step 19:** The AI-DAPT Platform stores the re-trained model. **Step 20:** The AI-DAPT User receives a notification about the action.

Scenario Workflow Diagram	

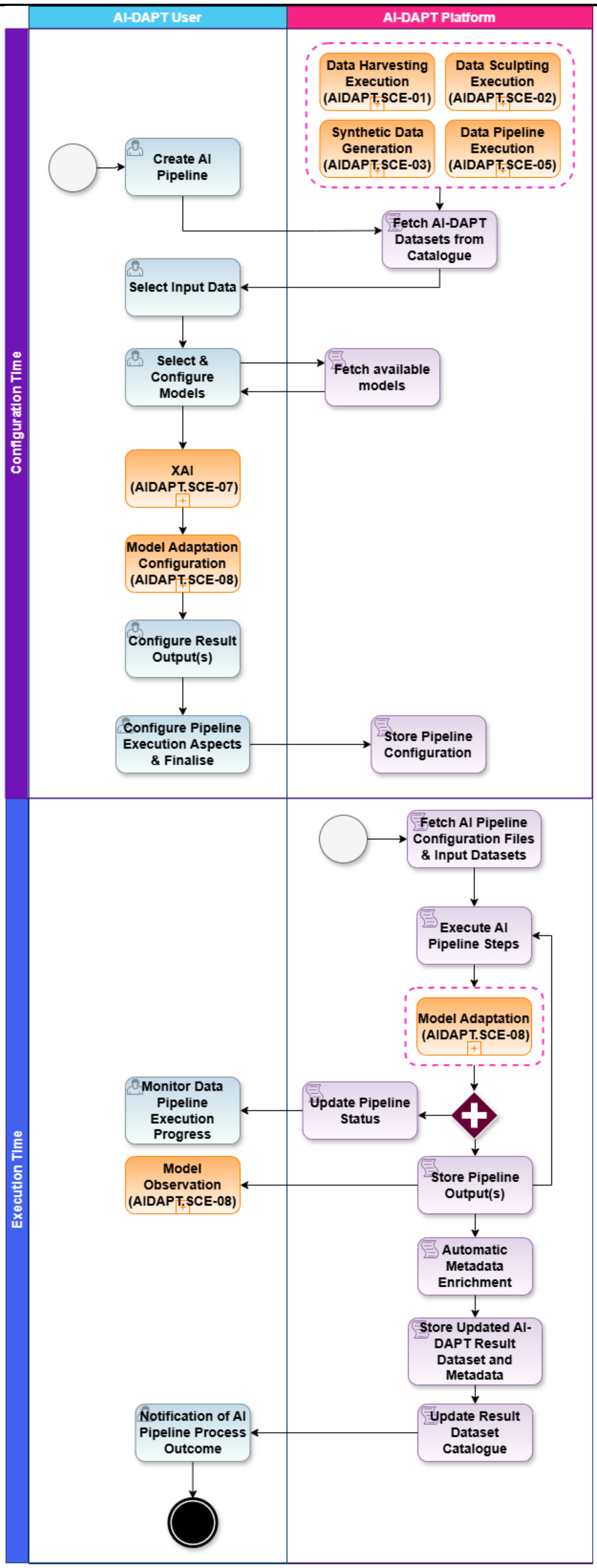
AI-DAPT User
AI-DAPT Platform
Configuration Time
Enter Model Adaptation
Fetch AI-DAPT Model Artefacts from Catalogue
Select Model Artefact for Application of Adaptation Rules
Temporal-based Rules
Quality-based Rules
Define Temporal Condition
Define Observability Metric Condition
Define Action of Adaptation Rule
Save Adaptation Rule
Store Adaptation Rule
Define more rules?
YES
NO
Finalise Adaptation Configuration
Store Configuration
Execution Time
Fetch Adaptation Configuration
Fetch Model Results
Calculate & Store Observability Metrics
Check Quality-based Rules
Quality Violation?
YES
Apply Rule Actions & Save Changes
Alert for Quality Violation & Actions Taken
Inspect Observability Metrics
Request Comparative Analysis
Try alternative Algorithms
Compare Results
Alternative has better Performance?
YES
Recommendation for Best-fit Alternative
Accept Changes?
YES
Apply & Store Changes
Re-execute with the Updated Configuration & Store Results
Notification about Applied Changes

	 AI-DAPT User AI-DAPT Platform Execution of Temporal Rules Triggered due to Temporal Adaptation Rule Fetch Model Fetch latest Data Retrain Model Store retrained Model Notification of Re-training Action and Outcome
Users Involved	Business User (possibly limited expertise for advanced configurations), Data Scientist
Users' Benefits	Users ensure the high performance of their models during the production phase, taking advantage of continuous learning and adaptive AI techniques. They are provided with the appropriate observability metrics, alerts and recommendations, while they can also define their own thresholds and adaptation rules. The modular design of the adaptation services allows their infusion in AI Pipelines, in a flexible manner.
Challenges	Interplay of adaptation with AI Pipeline Execution Engine. Design of case-agnostic/generic observability framework. Design of efficient model adaptation triggering mechanisms (e.g., avoiding unnecessary model re-trainings)
Success Criteria	The calculated observability metrics for an AI-DAPT Model results are available to the AI-DAPT User. Alerts are generated according to user preferences. Model adaptation suggestions are provided by the AI-DAPT Platform. AI Pipelines/Services are adapted and re-executed according to adaptation configuration.

5.2.9 SCE-09: Hybrid AI Pipeline Design and Execution

Hybrid AI Pipeline Design and Execution - [SCE-09]	
Scenario ID	SCE-09
Scenario Name	AI Pipeline Design and Execution
Scenario Overview	The authenticated AI-DAPT User wants to create a Hybrid AI Pipeline that will be executed based on their specific needs. For this purpose, the AI-DAPT User creates a custom AI Pipeline, defines its execution specificities, the AI-DAPT dataset that will be used, selects the ML and AI steps they want to be applied and configures their parameters. Furthermore, the AI-DAPT User can infuse in the Pipeline also XAI and Adaptation features. Once ready, the AI-DAPT User finalises the pipeline configuration. The AI Pipeline is executed according to the configuration, with intermediate results stored and annotated to allow the execution of the next step of the pipeline, resulting in the final AI-DAPT result dataset that is persisted in the AI-DAPT Scalable Storage Services, along with new metadata generated by the Documentation Engine. If adaptation has been infused in the pipeline, the appropriate adaptations are made to the pipeline at production time. The AI-DAPT User has an overview of the AI Pipeline Execution status per step and receives the relevant notification upon finalisation of the execution.
Scenario Steps	**Configuration Time** **Step 1:** The AI-DAPT User creates a new AI Pipeline. **Step 2:** The AI-DAPT User selects the AI-DAPT datasets that will be utilised (coming from execution of SCE-1 and optionally SCE-2, OR SCE-03, OR Data Pipeline (SCE-05). **Step 3:** The AI-DAPT User selects the models that should be used from the ones available and configures their parameters. **Step 4:** The AI-DAPT User infuses XAI techniques (SCE-07) in the AI Pipeline. **Step 5:** The AI-DAPT User adds and defines observability and adaptation steps in the AI-Pipeline (SCE-08). **Step 6:** The AI-DAPT User defines the AI Pipeline output(s) configuration for the storage of the pipeline result(s) in production. **Step 7:** The AI-DAPT User reviews the configured AI Pipeline, configures its execution aspects (e.g. periodicity) and finalises the configuration. **Step 8:** The AI-DAPT Platform stores the AI Pipeline configuration and trained models. **Execution Time** **Step 9:** The AI-DAPT Platform is triggered to initiate the execution of the AI Pipeline according to the AI Pipeline execution configuration. **Step 10:** The AI-DAPT Platform retrieves the configuration files of the AI Pipeline, the input AI-DAPT datasets, and the models **Step 11:** The AI-DAPT Platform executes the AI Pipeline steps.

	Step 12: The AI-DAPT Platform attaches additional metadata to the generated result(s). **Step 13:** The AI-DAPT Platform stores the AI-DAPT result dataset(s) and metadata. **Step 14:** The result dataset(s) are available for use in other AI-DAPT Services and in AI Pipelines. **Step 15:** The AI-DAPT User monitors and receives notifications about the completion of the execution of the AI Pipeline, while they can also view the observability metrics if available (SCE-08).

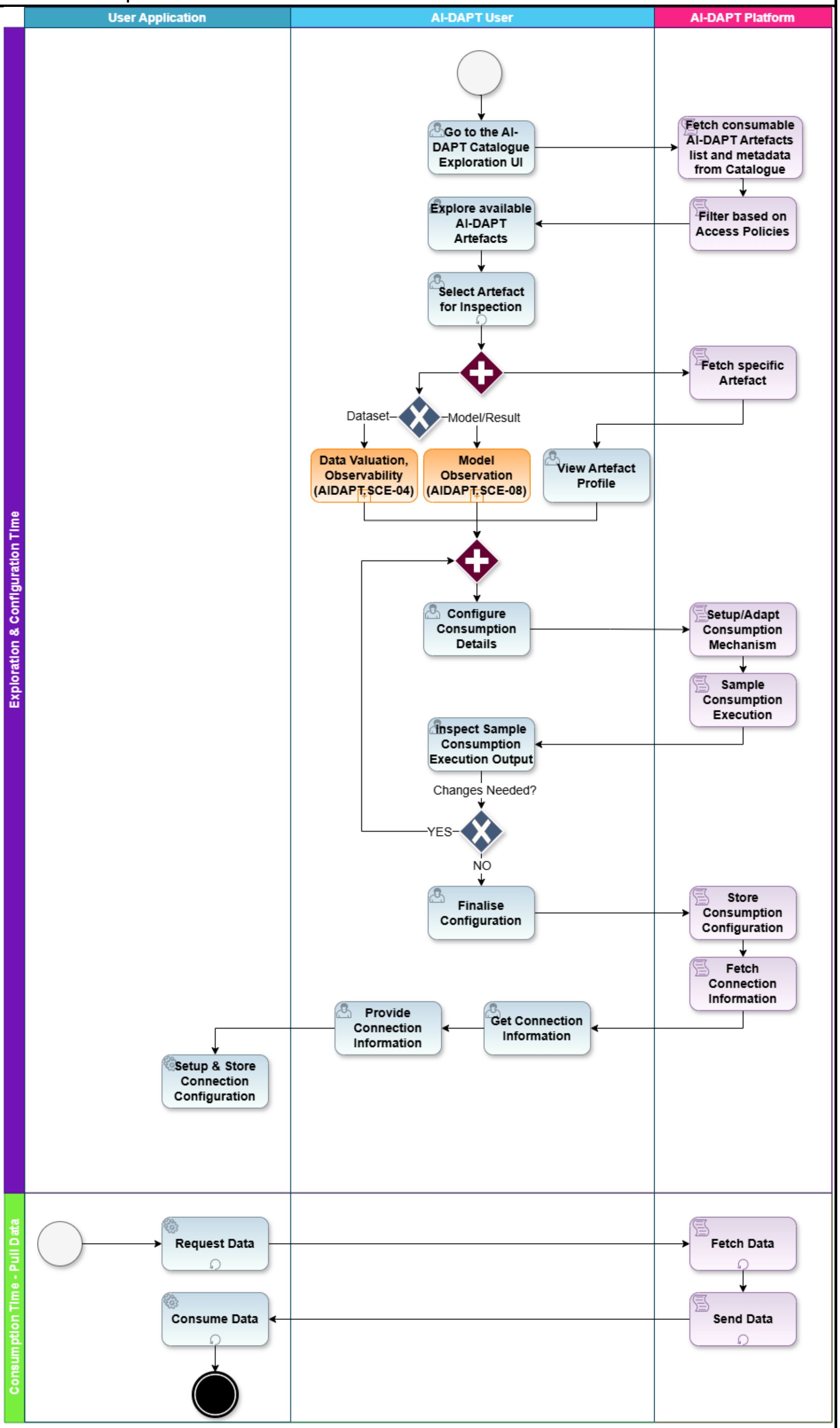
Scenario Workflow Diagram	

AI-DAPT User
AI-DAPT Platform
Configuration Time
Execution Time
Data Harvesting Execution (AIDAPT.SCE-01)
Data Sculpting Execution (AIDAPT.SCE-02)
Synthetic Data Generation (AIDAPT.SCE-03)
Data Pipeline Execution (AIDAPT.SCE-05)
Create AI Pipeline
Fetch AI-DAPT Datasets from Catalogue
Select Input Data
Select & Configure Models
Fetch available models
XAI (AIDAPT.SCE-07)
Model Adaptation Configuration (AIDAPT.SCE-08)
Configure Result Output(s)
Configure Pipeline Execution Aspects & Finalise
Store Pipeline Configuration
Fetch AI Pipeline Configuration Files & Input Datasets
Execute AI Pipeline Steps
Model Adaptation (AIDAPT.SCE-08)
Monitor Data Pipeline Execution Progress
Update Pipeline Status
Model Observation (AIDAPT.SCE-08)
Store Pipeline Output(s)
Automatic Metadata Enrichment
Store Updated AI-DAPT Result Dataset and Metadata
Notification of AI Pipeline Process Outcome
Update Result Dataset Catalogue

Users Involved	Business User (possibly limited expertise for advanced configurations), Data Scientist
Users' Benefits	Users can create custom AI Pipelines utilising in a flexible manner the AI-DAPT AI services (default and custom models, experimentation, XAI, observability and adaptation). The Users stay in control of their pipelines at production and can monitor closely both the model performance as well as the pipeline status per se.
Challenges	Interplay of the self-standing AI experimentation, design and monitoring/adaptation services with Pipeline Designer and Execution Engine. Design of case-agnostic/generic hybrid AI models. Translation of complex AI services in AI Pipeline blocks/operators.
Success Criteria	AI operations results are generated by the AI-DAPT Platform according to the defined AI Pipelines. The outputs of an AI Pipeline are available for further use within the AI-DAPT Platform.

5.2.10 SCE-10: AI-DAPT Datasets/Results Consumption

AI-DAPT Datasets/Results Consumption - [SCE-10]	
Scenario ID	SCE-10
Scenario Name	AI-DAPT Datasets/Results Consumption
Scenario Overview	The authenticated AI-DAPT User wants to utilise the outputs of the AI-DAPT Pipelines and services in their applications and systems. The AI-DAPT User explores the available consumable artefacts - i.e. datasets (outputs of Data Pipelines and data-related services) and results and models (outputs of AI-Pipelines and AI-related services)) in the AI-DAPT catalogue. Apart from standard metadata and a sneak-peak inside the artefact (e.g. included fields and structure), the AI-DAPT User can inspect the valuation and observability metrics of each AI-DAPT consumable artefact if available. The AI-DAPT User selects the artefacts they want to consume in their applications, and configures the consumption aspects. The AI-DAPT Platform stores the configuration, in case the AI-DAPT user wants to revisit it in the future, sets up the appropriate underlying retrieval mechanism, generates any required information for data consumption (e.g. API call). The AI-DAPT User can afterwards utilise this information in their applications in order to retrieve the data stored in the AI-DAPT artefact(s).
Scenario Steps	**Exploration & Configuration Time** **Step 1:** The AI-DAPT User explores the consumable artefacts of the AI-DAPT Catalogue utilising the available exploration filters. **Step 2:** The AI-DAPT User selects an artefact to view more information. **Step 3:** The AI-DAPT platform fetches the artefact stored information. **Step 4:** The AI-DAPT User inspects the artefact metadata and preview information. **Step 5:** The AI-DAPT User inspects the artefact valuation and observability metrics if available (SCE-04). **Step 6:** The AI-DAPT User configures the consumption aspects (e.g. mode, selection of subset of available fields in the dataset etc). **Step 7:** The AI-DAPT Platform sets up the consumption mechanism according to user selections. **Step 8:** The AI-DAPT Platform creates a sample preview of the dataset that will be consumed based on user selections. **Step 9:** The AI-DAPT User inspects the sample and makes changes to the configuration if needed. **Step 10:** Repeat steps 7 – 9 until satisfied with the result. **Step 11:** The AI-DAPT User finalises the configuration. **Step 12:** The AI-DAPT Platform stores the configuration and provides the AI-DAPT User with the connection details. **Step 13:** The AI-DAPT User utilises the provided connection information in their application to setup the connection.

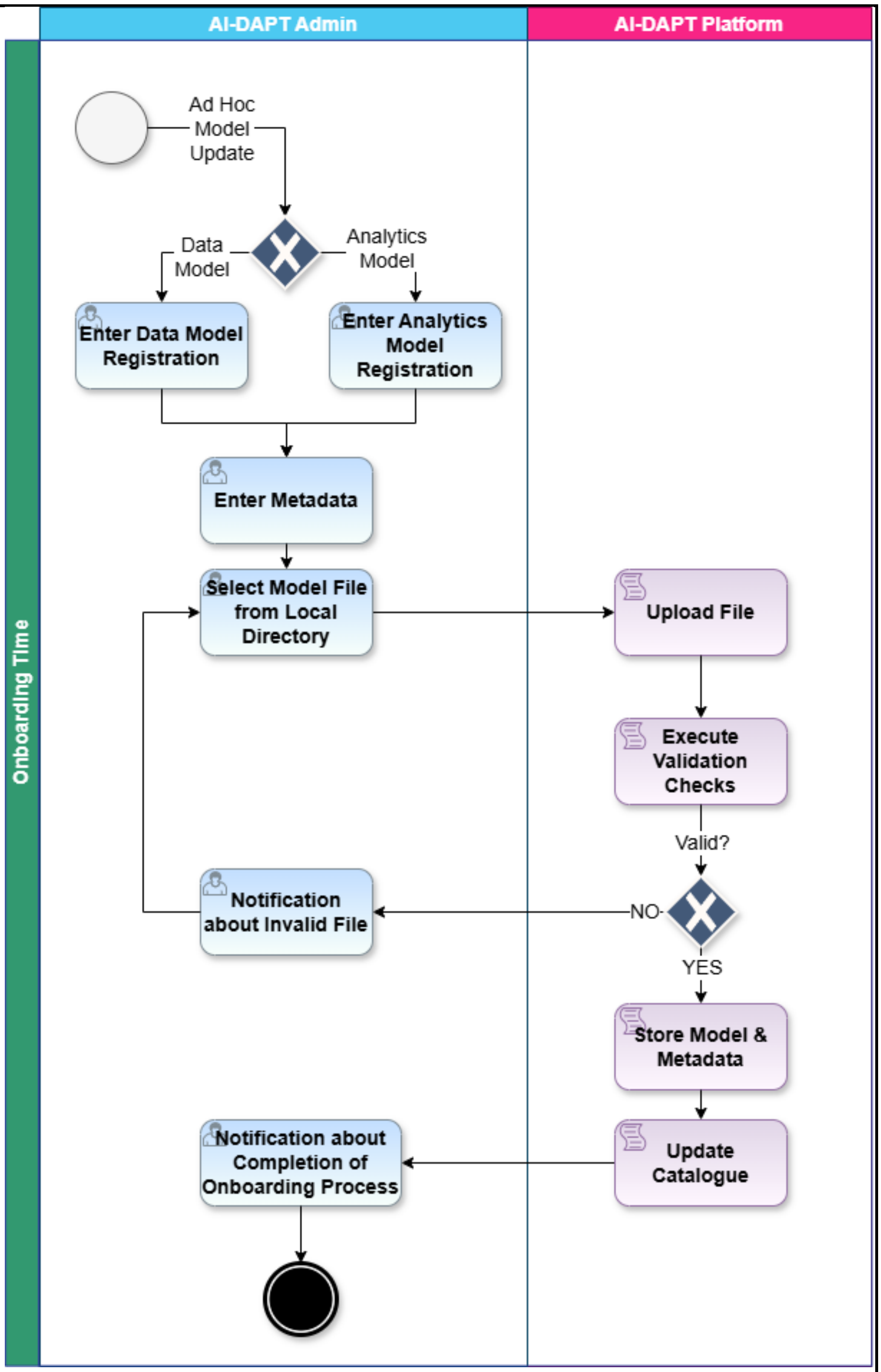
	Consumption Time – Pull Data **Step 14:** The Application requests data from the AI-DAPT Platform according to the application needs (e.g. one request every 5 minutes). **Step 15:** The AI-DAPT Platform receives the request, fetches the data and responds.
Scenario Workflow Diagram	
Users Involved	Business User, Data Scientist

User Application
AI-DAPT User
AI-DAPT Platform
Exploration & Configuration Time
Consumption Time - Pull Data
Go to the AI-DAPT Catalogue Exploration UI
Fetch consumable AI-DAPT Artefacts list and metadata from Catalogue
Explore available AI-DAPT Artefacts
Filter based on Access Policies
Select Artefact for Inspection
Fetch specific Artefact
Dataset
Model/Result
Data Valuation, Observability (AIDAPT.SCE-04)
Model Observation (AIDAPT.SCE-08)
View Artefact Profile
Configure Consumption Details
Setup/Adapt Consumption Mechanism
Sample Consumption Execution
Inspect Sample Consumption Execution Output
Changes Needed?
YES
NO
Finalise Configuration
Store Consumption Configuration
Fetch Connection Information
Get Connection Information
Provide Connection Information
Setup & Store Connection Configuration
Request Data
Fetch Data
Send Data
Consume Data

Users' Benefits	Users can retrieve the artefacts created in AI-DAPT (i.e. datasets, results, models) to use in their own applications, utilising flexible retrieval mechanisms.
Challenges	Generic/application-agnostic exposure mechanisms might not be able to support the specificities of every possible external application that wants to consume data from AI-DAPT. Access and retrieval control should be applied.
Success Criteria	The AI-DAPT User can find all available consumable artefacts (AI-DAPT datasets, results, models) they are eligible to. An external application consistently consumes data from the AI-DAPT Platform according to configuration.

5.2.11 SCE-11: Data and Analytics Models Registration

Data and Analytics Models Registration - [SCE-11]	
Scenario ID	SCE-11
Scenario Name	Data and Analytics Models Registration
Scenario Overview	The authenticated AI-DAPT User wants to onboard to the AI-DAPT Platform data models and analytics models that will be available for use in the semantic reconciliation of their data and their AI operations respectively. The AI-DAPT User uploads the model file and defines its metadata through the AI-DAPT Platform. The AI-DAPT Platform validates the file, and if everything is ok, the model is stored and available to be utilised. The AI-DAPT User can define whether the onboarded model is a totally new model, or a new version of an existing model.
Scenario Steps	**Onboarding Time** **Step 1:** The AI-DAPT User selects whether the onboarding action concerns Data Model or Analytics Model onboarding. **Step 2:** The AI-DAPT User selects from their local directory the relevant file. **Step 3:** The AI-DAPT User defines the model metadata. **Step 4:** The AI-DAPT Platform gets the file and performs the relevant validation checks. **Step 5:** If the file is valid and compliant with the AI-DAPT model structure guidelines, the AI-DAPT Platform stores the file and the relevant metadata. **Step 6:** If it is not valid, the AI-DAPT User receives a notification and needs to upload the corrected file. Repeat steps 2 – 5 until there are no validation errors. **Step 7:** The AI-DAPT Platform updates the AI-DAPT Catalogue, and the new model is available for the AI-DAPT Users to utilise. **Step 8:** The AI-DAPT User receives a notification about the outcome of the onboarding process.

Scenario Workflow Diagram	 AI-DAPT Admin AI-DAPT Platform Onboarding Time Ad Hoc Model Update Data Model Analytics Model Enter Data Model Registration Enter Analytics Model Registration Enter Metadata Select Model File from Local Directory Upload File Execute Validation Checks Valid? NO YES Notification about Invalid File Store Model & Metadata Update Catalogue Notification about Completion of Onboarding Process
Users Involved	Business User, Data Analyst
Users' Benefits	End-users can easily extend and maintain the built-in data and analytics model library of AI-DAPT and are indirectly benefited, as the new/updated models become available for usage in their data and AI operations afterwards through the AI-DAPT catalogue.
Challenges	Version control and backwards compatibility.
Success Criteria	The required AI-DAPT format for data and analytics models should be as close to standardised formats as possible.

6 Technical Requirements

This section outlines the technical requirements for the AI-DAPT platform, identified through iterative brainstorming sessions and analysis of user stories from D1.1 of WP1, as well as the end-to-end usage scenarios discussed in the previous section. Through repeated collaboration with consortium partners, requirements were mapped to specific tasks and components, with each partner refining suggestions and addressing any gaps. This collaborative approach ensured that the requirements addressed both the functional needs of the users and the technical constraints of the platform, resulting in a first draft and actionable set of technical specifications for the AI-DAPT MVP detailed in the Section 7.

The extraction of technical requirements followed an agile methodology, with a primary focus on user stories developed collaboratively between demonstrator's (end-users) and technical partners, as documented in the previous D1.1 of WP1. The agile process began with the definition of the user stories, identified from the needs of the demonstrators and from general usage scenarios for the AI-DAPT platform. Refining the platform's functional and non-functional requirements will be an ongoing process until the end of the project.

The technical requirements for AI-DAPT are divided into two main categories:

- **Functional Requirements:** They define the core features, services and functionalities that the AI-DAPT platform shall provide to its users. Additionally, they specify the requirements needed to address the identified needs, as determined by the recorded user stories and usage scenarios.
- **Non-functional Requirements**: These focus on the quality attributes the system must exhibit to effectively support the functional requirements, such as performance, scalability, reliability, usability, and security. They ensure that the platform not only functions as intended but also works as expected whilst meeting user expectations under varying conditions.

Each technical requirement is assigned a unique ID, a short description, and is mapped to the related user stories and usage scenarios. Additionally, each requirement is categorized and prioritized on a scale from 1 to 4, based on its importance to the platform's functionality:

1 - *Critical*

2 - *High*

3 - *Medium*

4 - *Low*

This prioritization reflects the necessity of each requirement for the platform's effective operation in alignment with the AI-DAPT's principles. To evaluate the prioritization of each technical requirement, we analyse the relevant demonstrator user needs and usage scenarios. The priority of each user need was identified through interviews with demonstrators (from D1.1), which guide the critical level each requirement should have. Moreover, the usage scenarios were created with the collaboration of the technical partners, defining the platform's core functionality. Therefore, we employ a scoring system that combines them, to ensure each requirement gets the corresponding priority. A requirement adopts the priority of each relevant user need (1-3 range), as well as 3 points when it is part of the platform's core functionality. Figure 6.1.1 shows the evaluation for the technical requirements and Figure 6.1.2 shows the evaluation for the non-technical requirements.

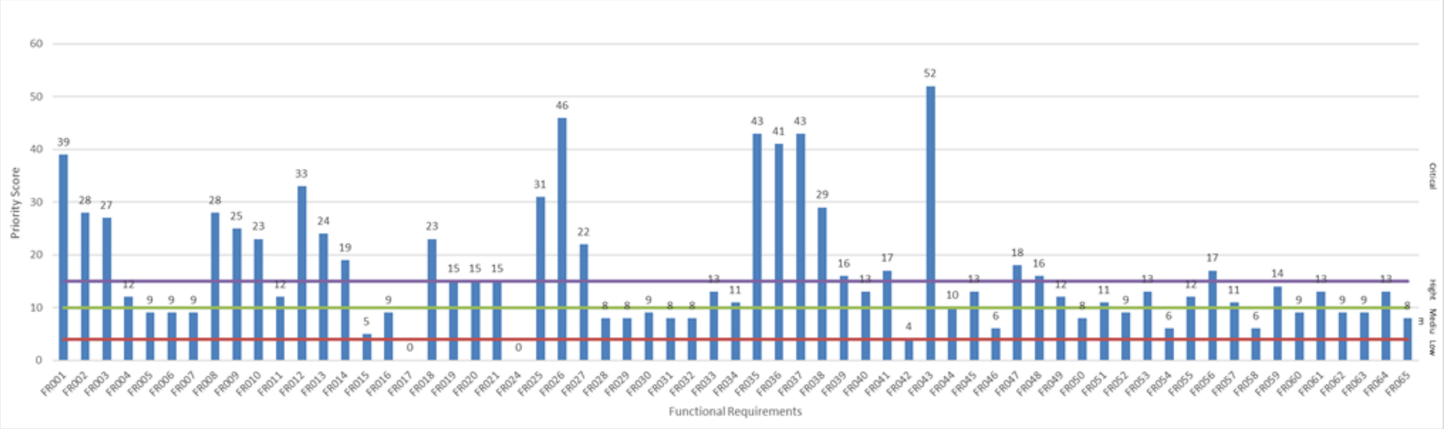


Figure 6.1.1: Priority Score Based on User Stories & Usage Scenarios for Functional Requirements.

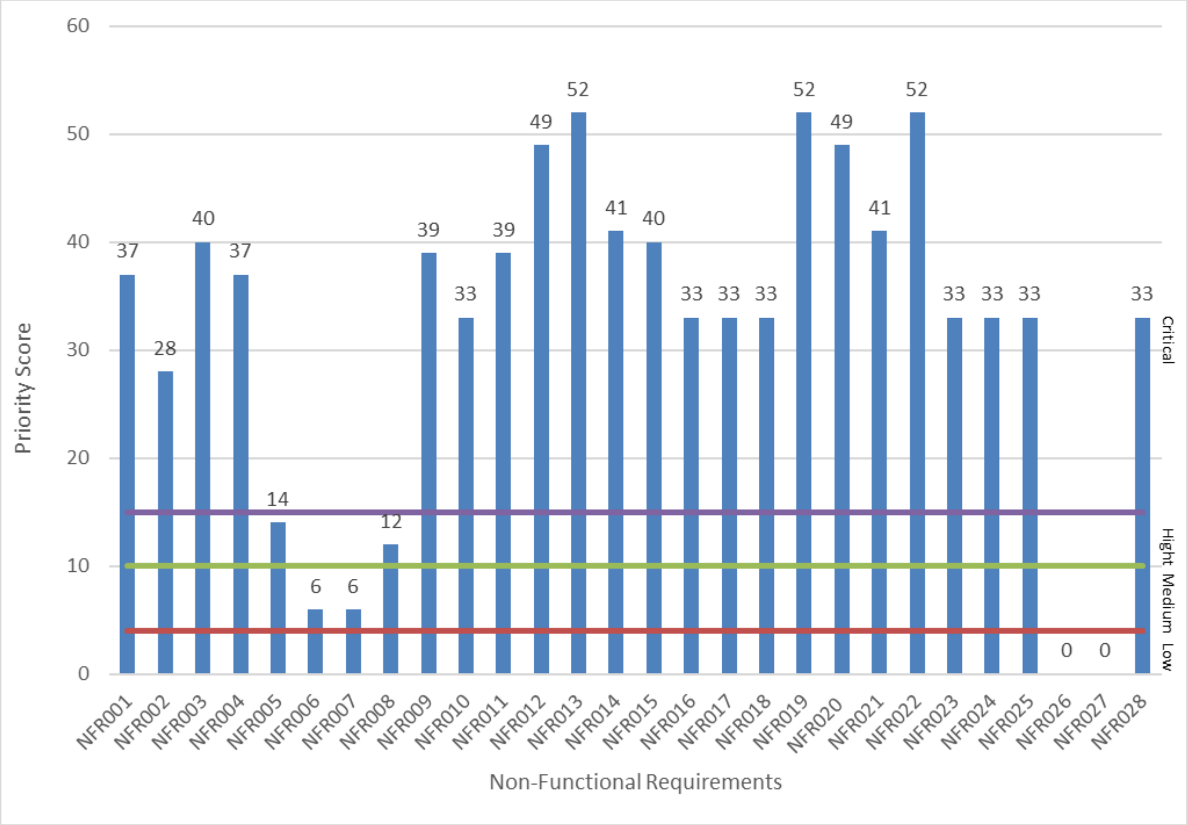


Figure 6.1.2: Priority Score Based on User Stories & User Scenarios for Non-Functional Requirements.

6.1 Functional Requirements

This subsection lists all the functional requirements that the AI-DAPT platform has to address in order to meet both user needs and usage scenarios. These requirements are organized under five categories: AI Pipeline & Analytics, Data Collection & Exploration, Data Privacy, Notification, and Other. The priority of each requirement is identified from the methodology explained above, for which the results are shown in Figure 7.1.1.

AI Pipeline & Analytics			
ID	**Description**	**Related User Story / Usage Scenario**	**Priority**
FR002	AI-DAPT should provide users with the option to select and apply algorithms from a predefined list.	SCE-06, SCE-07, SCE-09, D2_US1, D2_US11, D3_US1, D3_US2, D4_US16, D4_US2, D4_US19	1- Critical
FR004	AI-DAPT should allow users to create and experiment with algorithms through the use of Jupyter notebooks.	SCE-06, D1_US3, D2_US9, D4_US2, D3_US3, D3_US7,D3_US8,D3_US11	2- High
FR005	AI-DAPT should allow users to upload and download algorithms.	SCE-06, D1_US3, D2_US9, D4_US2, D3_US7,D3_US8,D3_US11	3- Medium
FR006	AI-DAPT should enable users to save and reuse algorithms in their AI pipelines.	SCE-06, D1_US3, D2_US9, D4_US2, D3_US7,D3_US8,D3_US11	3- Medium
FR009	AI-DAPT should allow the users to monitor the progress of their AI pipelines in real-time.	SCE-08, D1_US7, D1_US15, D1_US14, D1_US16, D2_US2, D2_US6, D2_US8, D4_US24, D1_US3, D1_US8	1- Critical
FR013	AI-DAPT should allow the user to understand the outcomes of their AI pipelines by utilising XAI techniques and visualisations.	SCE-07, D1_US3, D1_US8, D3_US2, D3_US3, D4_US16, D4_US2, D4_US19, D1_US6, D1_US10	1- Critical
FR014	AI-DAPT should suggest to its users data operations that can optimize their models and pipelines.	SCE-07, D1_US3, D1_US8, D3_US3, D4_US16, D4_US2, D4_US19, D1_US10	1- Critical
FR015	AI-DAPT should provide the ability to compare the results and/or performance of different AI models from previous trials.	SCE-09, D3_US10, D3_US11	3- Medium
FR020	AI-DAPT should provide pre-built templates or workflows for the AI pipelines.	SCE-05, SCE-09, D1_US9, D1_US3, D1_US8, D3_US10, D3_US11	2- High
FR021	AI-DAPT should provide tools for configuring AI model training & inference parameters.	SCE-05, SCE-09, D1_US9, D1_US3, D1_US8, D3_US10, D3_US11	2- High
FR025	AI-DAPT should support real-time training & inference of AI models.	SCE-06, SCE-09, D1_US9, D1_US13, D2_US1, D2_US11, D3_US1, D3_US2, D4_US16, D4_US2, D4_US19	1- Critical
FR026	AI-DAPT should support fundamental AI methods and techniques, including bias detection and data valuation.	SCE-02, SCE-03, SCE-04, SCE-06, D1_US3, D1_US8, D2_US1, D2_US11, D3_US1, D3_US2,	1- Critical

		D3_US3, D4_US16, D4_US2, D4_US19, D3_US7, D3_US8, D3_US9, D3_US10, D3_US11, D4_US27	
FR036	AI-DAPT should support the execution of AI pipelines using as input one or more (multiple) datasets.	SCE-02, SCE-03, SCE-06, D1_US3, D1_US8, D2_US1, D2_US11, D3_US1, D3_US2, D3_US3, D4_US16, D4_US2, D4_US19, D3_US7, D3_US8, D3_US10, D3_US11	1- Critical
FR037	AI-DAPT should enable a human-in-the-loop approach for critical tasks (e.g. data cleaning, synthetic data generation), allowing users to review, approve or modify changes before they are applied.	SCE-02, SCE-03, SCE-06, D1_US3, D1_US8, D2_US1, D2_US11, D3_US1, D3_US2, D3_US3, D4_US16, D4_US2, D4_US19, D3_US7, D3_US8, D3_US9, D3_US10, D3_US11, D4_US27	1- Critical
FR048	AI-DAPT should allow integration of hybrid models by combining ML & DL algorithms with science-based (first-principles) models.	SCE-09, D3_US10, D4_US3, D3_US11, D1_US3, D3_US3, D4_US16	1- Critical
FR049	AI-DAPT should enable hyperparameter tuning within all models' definitions (ML, DL & hybrid).	SCE-06, D3_US11, D1_US8, D3_US10, D4_US7, D1_US6	2- High
FR050	AI-DAPT should enable comparison of AI model results and performance from different trials.	SCE-08, D1_US6, D1_US3, D3_US11	3- Medium
FR051	AI-DAPT should support adaptive AI techniques for evaluating and optimizing for both pre-defined (e.g. holdout, cross-validation, accuracy, etc) and user-defined metrics	SCE-08, D1_US3, D1_US8, D3_US11, D3_US3	2- High
FR052	AI-DAPT should offer version control and management for AI models, facilitating rollback, comparison, and deployment of different model versions.	SCE-09, D1_US3, D3_US10, D3_US11, D3_US8	3- Medium
FR056	AI-DAPT should monitor data at every stage of the AI pipeline to identify potential data issues.	SCE-08, D1_US15, D1_US14, D1_US16, D2_US4, D2_US2, D2_US8, D4_US24	1- Critical
FR057	AI-DAPT should allow users to select and define the necessary metrics used for evaluating their models.	SCE-04, D2_US4, D1_US3, D3_US10, D4_US3, D3_US11	2- High

Data Collection & Exploration			
ID	Description	Related User Story / Usage Scenario	Priority
FR008	AI-DAPT should support connection to various APIs for data import from other platforms, devices and services.	SCE-01, D1_US12, D4_US3, D4_US25, D1_US7, D1_US15, D1_US14, D1_US16, D2_US2, D2_US6, D2_US8	1- Critical
FR012	AI-DAPT should offer customizable data visualisation tools (e.g. dashboards, graphs) to help users explore data insights and AI model decisions.	SCE-04, D1_US7, D1_US15, D1_US14, D1_US16, D2_US2, D2_US6, D2_US8, D4_US5, D4_US6, D4_US8, D4_US9, D4_US10, D4_US15	1- Critical
FR016	AI-DAPT should provide capabilities to add additional semantic information (metadata) to datasets.	SCE-01, SCE-02, D1_US9	3- Medium
FR017	AI-DAPT should allow datasets and metadata to be catalogued, providing information for evaluating the fitness of the data for its intended uses.	-	4- Low
FR028	AI-DAPT should support both automatic and manual data annotation to common information data models.	SCE-02, D1_US3, D3_US2	3- Medium
FR029	AI-DAPT should support full synthetic data generation.	SCE-03, D4_US27, D3_US9, D4_US26	3- Medium
FR030	AI-DAPT should support data generation from scratch and from existing datasets.	SCE-03, D3_US9, D4_US27, D4_US26, D2_US4	3- Medium
FR031	AI-DAPT should support multiple data sampling methods (e.g. random sampling, specific interval sampling).	SCE-02, D3_US7, D3_US8, D3_US9	3- Medium
FR032	AI-DAPT should provide feature selection capabilities using filter, wrapper and embedded methods to optimize the feature set for ML models.	SCE-02, D3_US10, D3_US11, D3_US2	3- Medium
FR033	AI-DAPT should offer feature engineering tools, allowing for the handling of missing data (using imputation method), categorical encoding, variable transformation, discretization,	SCE-02, D3_US9, D4_US26, D4_US27, D4_US2, D4_US5	2- High

	outlier engineering, and date/time feature engineering. Users should be able to configure or select the methods used for each task.		
FR034	AI-DAPT should offer data cleaning mechanisms, such as handling missing values, removing duplicates, correcting data inconsistencies and normalizing data formats.	SCE-02, D4_US27, D2_US4, D1_US13, D1_US8, D3_US9	2- High
FR035	AI-DAPT should support the use of machine learning techniques, whilst keeping a human-in-the-loop, for data annotation.	SCE-02, SCE-03, SCE-06, D1_US3, D1_US8, D2_US1, D2_US11, D3_US1, D3_US2, D3_US3, D4_US16, D4_US2, D4_US19, D3_US7, D3_US8, D3_US9, D3_US10, D3_US11, D4_US27	1- Critical
FR038	AI-DAPT should support file upload, download and conversion services for common text formats (e.g. CSV, JSON, TXT).	SCE-01, SCE-10, D4_US25, D4_US13, D4_US15, D4_US6, D4_US9, D1_US1, D1_US12, D2_US4, D4_US3	1- Critical
FR039	AI-DAPT should allow users to create new versions of their datasets.	SCE-05, SCE-10, D1_US13, D4_US26, D1_US10, D3_US9, D4_US27	1- Critical
FR040	AI-DAPT should support tags (e.g. real-time data, historical, synthetic, owned) & categories (e.g. operational, medicine) for datasets.	SCE-01, SCE-02, SCE-03, D1_US7, D3_US9	2- High
FR041	AI-DAPT should support both time-series and non-time series structured data.	SCE-09, D1_US8, D3_US2, D4_US16, D3_US7, D4_US10, D4_US19	1- Critical
FR042	AI-DAPT should provide baseline open-datasets that users can utilize as examples for their pipelines.	SCE-10, D3_US9	4- Low
FR044	AI-DAPT should support search functionality for datasets and algorithms, allowing users to find relevant assets based on their name, category, tag, keyword, date and time.	SCE-10, D4_US9, D2_US2, D2_US6, D2_US8	3- Medium

FR045	AI-DAPT should be able to combine multiple datasets into a single, newly created dataset.	SCE-05, SCE-09, D4_US26, D1_US10, D3_US9, D4_US27	2- High
FR046	AI-DAPT should maintain a history log for dataset actions, tracking events such as when the dataset was created, updated or modified.	SCE-10, D4_US26	3- Medium
FR053	AI-DAPT should support data uploads via batch files and datasets.	SCE-01, D1_US1, D1_US12, D2_US4, D4_US25	2- High
FR054	AI-DAPT should support the import of data through streaming mechanisms.	SCE-01, D1_US12	3- Medium
FR055	AI-DAPT should validate that incoming data is in the correct format.	SCE-01, D4_US26, D3_US9, D1_US8, D4_US25	2- High
FR058	AI-DAPT should allow users to export their data and AI pipeline configurations for import into external AI-DAPT compliant AI solutions.	SCE-10, D4_US3	3- Medium
FR060	AI-DAPT should provide the capability to search the semantic information (metadata) of datasets.	SCE-01, SCE-02, D1_US9	3- Medium
FR061	AI-DAPT should automatically map raw data formats to standard data models using semantic concepts and entities.	SCE-01, SCE-02, D2_US4, D1_US3, D1_US8, D1_US3	2- High
FR062	AI-DAPT should allow users to explore datasets through the use of Jupyter notebooks.	SCE-02, D1_US3, D1_US8, D2_US4, D3_US10	3- Medium
FR063	AI-DAPT should allow users to persist/store their experimentations in the form of Jupyter notebooks.	SCE-06, D1_US3, D1_US8, D2_US4, D4_US27	3- Medium
FR064	AI-DAPT should allow users to visualize algorithms and datasets through the use of Jupyter notebooks.	SCE-02, D1_US3, D1_US6, D2_US4, D3_US8, D4_US16	2- High
FR065	AI-DAPT should support partially synthetic data generation.	SCE-03, D4_US27, D3_US9, D4_US26	3- Medium

Data Privacy			
ID	**Description**	**Related User Story / Usage Scenario**	**Priority**
FR027	AI-DAPT should allow the user perform data anonymization over their datasets.	SCE-03, D1_US2, D1_US5, D2_US3, D2_US10, D2_US7, D3_US5, D4_US1	1- Critical
FR043	AI-DAPT should apply multiple levels of data confidentiality to control access to datasets by the users. Each user must be assigned to a different role, which will determine their ability to upload, view, modify and export datasets.	SCE-01, SCE-02, SCE-03, SCE-04, SCE-05, SCE-06, SCE-07, SCE-08, SCE-09, SCE-10, SCE-11, D1_US2, D1_US5, D2_US3, D2_US10, D2_US7, D3_US5, D4_US1	1- Critical

Notification			
ID	**Description**	**Related User Story / Usage Scenario**	**Priority**
FR010	AI-DAPT should provide users with real-time notifications about dataset status updates, including changes to dataset availability, modifications to dataset content, processing status, and any errors that occur during dataset operations, as well as updates regarding AI/Data pipeline stages.	SCE-04, SCE-08, D4_US23, D4_US17, D1_US6, D1_US2, D4_US11, D4_US14	1- Critical
FR011	AI-DAPT should inform the user on optimizations, like applying Adaptive AI techniques, that can happen during the execution of AI pipelines through real-time notifications.	SCE-04, SCE-08, D4_US17, D4_US14	2- High
FR047	AI-DAPT should allow users to customize their notification preferences, such as selecting the types of notifications they wish to receive, the frequency of notifications and their preferred delivery method (e.g. through the platform, email or both).	SCE-04, SCE-08, D4_US23, D4_US17, D4_US11, D4_US14	1- Critical
FR059	AI-DAPT should notify users when their models fail to meet performance thresholds or exhibit suboptimal performance.	SCE-08, D2_US4, D1_US3, D1_US8, D4_US14, D4_US23	2- High

Other			
ID	Description	Related User Story / Usage Scenario	Priority
FR001	AI-DAPT should provide appropriate logging mechanisms for all operations/components.	SCE-01, SCE-02, SCE-03, SCE-04, SCE-05, SCE-06, SCE-07, SCE-08, SCE-09, SCE-10, SCE-11, D1_US2, D2_US10	1- Critical
FR003	AI-DAPT should allow users to schedule operations for automated execution.	SCE-05, D1_US3, D1_US4, D1_US8, D2_US4, D2_US9, D2_US11, D3_US6, D4_US7, D4_US11, D4_US14	1- Critical
FR007	AI-DAPT shall support integration with existing software platforms relevant to the users' pipelines.	SCE-10, D4_US3, D1_US1	3- Medium
FR018	AI-DAPT should handle and be able to store large datasets.	SCE-01, D1_US1, D1_US7, D1_US8, D3_US5, D4_US25, D4_US26, D4_US3	1- Critical
FR019	AI-DAPT should provide functions to create and manage shortcuts and workflows.	SCE-05, SCE-09, D2_US2, D2_US6, D2_US8, D4_US24	2- High
FR024	AI-DAPT should dynamically scale computational resources based on pipeline demands to optimize infrastructure usage.	-	4- Low

6.2 Non-Functional Requirements

In this subsection the non-functional requirements are defined. They constitute the quality attributes necessary for making the platform reliable, sustainable and robust. These requirements address key aspects such as documentation, maintainability, performance, portability, privacy, reliability, scalability, security, and usability. Each of these categories play an important role in keeping the platform running smoothly, flexible, and of high quality. The priority of each requirement is identified from the methodology explained above, for which the results are shown in Figure 7.1.2.

ID	Description	Related User Story / Usage Scenario	Category	Priority
NFR001	AI-DAPT should provide detailed user manual and detailed documentation.	SCE-01, SCE-02, SCE-03, SCE-04, SCE-05, SCE-06, SCE-07, SCE-08, SCE-09, SCE-10, SCE-11, D1_US6, D2_US5	Documentation	1- Critical

NFR002	AI-DAPT must manage and store large datasets efficiently, with minimum processing time.	SCE-01, SCE-02, SCE-03, D4_US3, D4_US5, D4_US6, D4_US7, D4_US8, D4_US24, D4_US25, D4_US27	Performance	1- Critical
NFR003	AI-DAPT should ensure system components are implemented and operated independently, as to avoid code dependency, hence better maintainability and increased scalability.	SCE-01, SCE-02, SCE-03, SCE-04, SCE-05, SCE-06, SCE-07, SCE-08, SCE-09, SCE-10, SCE-11, D4_US15, D1_US1, D2_US4	Maintainability & Scalability	1- Critical
NFR004	AI-DAPT should support system upgrades with minimal downtime, ensuring that updates, patches, and new feature deployments can be implemented without significantly disrupting system operations.	SCE-01, SCE-02, SCE-03, SCE-04, SCE-05, SCE-06, SCE-07, SCE-08, SCE-09, SCE-10, SCE-11, D4_US9, D4_US10	Maintainability	1- Critical
NFR005	AI-DAPT should be able to execute AI pipelines and any kind of processing in a fast and efficient manner.	SCE-05, SCE-09, D2_US1, D4_US25, D3_US8, D1_US13	Performance	2- High
NFR006	AI-DAPT should support parallel processing to enhance the execution speed of AI operations.	SCE-06, SCE-09	Performance	3- Medium
NFR007	AI-DAPT must optimize its computational efficiency to handle complex algorithms without lag.	SCE-06, SCE-09	Performance	3- Medium
NFR008	AI-DAPT should effectively manage poor-quality, slow or hard-to-access data, finding optimal solutions to handle these scenarios while maintaining performance.	SCE-02, SCE-04, D4_US26, D2_US9, D4_US27	Performance	2- High
NFR009	AI-DAPT should be resilient, capable of handling software errors without impacting its overall functionality or availability.	SCE-01, SCE-02, SCE-03, SCE-04, SCE-05, SCE-06, SCE-07, SCE-08, SCE-09, SCE-10, SCE-11, D4_US26, D2_US9, D4_US27	Portability	1- Critical
NFR010	AI-DAPT should support deployment on various Linux distributions.	SCE-01, SCE-02, SCE-03, SCE-04, SCE-05, SCE-06, SCE-07, SCE-08, SCE-09, SCE-10, SCE-11	Portability	1- Critical
NFR011	AI-DAPT should offer an easy and straightforward installation process.	SCE-01, SCE-02, SCE-03, SCE-04, SCE-05, SCE-06, SCE-07, SCE-	Portability	1- Critical

		08, SCE-09, SCE-10, SCE-11, D4_US25, D4_US3		
NFR012	AI-DAPT must follow the relevant legal instruments, including GDPR (General Data Protection Regulation) and AI Act requirements and comply with data security policies.	SCE-01, SCE-02, SCE-03, SCE-04, SCE-05, SCE-06, SCE-07, SCE-08, SCE-09, SCE-10, SCE-11, D1_US2, D1_US5, D2_US10, D2_US7, D3_US5, D4_US1	Privacy	1- Critical
NFR013	AI-DAPT should ensure the privacy and protection of user data at all times.	SCE-01, SCE-02, SCE-03, SCE-04, SCE-05, SCE-06, SCE-07, SCE-08, SCE-09, SCE-10, SCE-11, D1_US2, D1_US5, D2_US3, D2_US10, D2_US7, D3_US5, D4_US1	Privacy	1- Critical
NFR014	AI-DAPT should support concurrent access for numerous users, maintaining high uptime for continuous access.	SCE-01, SCE-02, SCE-03, SCE-04, SCE-05, SCE-06, SCE-07, SCE-08, SCE-09, SCE-10, SCE-11, D4_US18, D3_US5, D2_US8	Reliability	1- Critical
NFR015	AI-DAPT should handle simultaneous requests and recover quickly from system failures to maintain reliability.	SCE-01, SCE-02, SCE-03, SCE-04, SCE-05, SCE-06, SCE-07, SCE-08, SCE-09, SCE-10, SCE-11, D4_US12, D2_US8, D3_US6	Reliability	1- Critical
NFR016	AI-DAPT must automatically back up data to prevent loss in case of system failures.	SCE-01, SCE-02, SCE-03, SCE-04, SCE-05, SCE-06, SCE-07, SCE-08, SCE-09, SCE-10, SCE-11	Reliability	1- Critical
NFR017	AI-DAPT should guarantee data consistency while applying data replication.	SCE-01, SCE-02, SCE-03, SCE-04, SCE-05, SCE-06, SCE-07, SCE-08, SCE-09, SCE-10, SCE-11	Reliability	1- Critical
NFR018	AI-DAPT should ensure all components have failover capabilities to maintain service continuity.	SCE-01, SCE-02, SCE-03, SCE-04, SCE-05, SCE-06, SCE-07, SCE-08, SCE-09, SCE-10, SCE-11	Reliability	1- Critical
NFR019	AI-DAPT should provide distinct public and private workspaces/databases to securely manage data access and support collaborative	SCE-01, SCE-02, SCE-03, SCE-04, SCE-05, SCE-06, SCE-07, SCE-08, SCE-09, SCE-10, SCE-11, D1_US2,	Security	1- Critical

	workflows. The platform should also support deployment on the user's private infrastructure for enhanced data control and security.	D1_US5, D2_US3, D2_US10, D2_US7, D3_US5, D4_US1		
NFR020	AI-DAPT should implement security measures, including end-to-end data encryption, to protect data during transmission and storage.	SCE-01, SCE-02, SCE-03, SCE-04, SCE-05, SCE-06, SCE-07, SCE-08, SCE-09, SCE-10, SCE-11, D1_US2, D1_US5, D2_US10, D2_US7, D3_US5, D4_US1	Security	1- Critical
NFR021	AI-DAPT should provide error protection methods for all input fields.	SCE-01, SCE-02, SCE-03, SCE-04, SCE-05, SCE-06, SCE-07, SCE-08, SCE-09, SCE-10, SCE-11, D2_US8, D3_US3, D4_US26, D1_US13	Security	1- Critical
NFR022	AI-DAPT should provide secure and controlled registration process for new users.	SCE-01, SCE-02, SCE-03, SCE-04, SCE-05, SCE-06, SCE-07, SCE-08, SCE-09, SCE-10, SCE-11, D1_US2, D1_US5, D2_US3, D2_US10, D2_US7, D3_US5, D4_US1	Security	1- Critical
NFR023	AI-DAPT should log all user actions to ensure traceability of operations.	SCE-01, SCE-02, SCE-03, SCE-04, SCE-05, SCE-06, SCE-07, SCE-08, SCE-09, SCE-10, SCE-11	Security	1- Critical
NFR024	AI-DAPT should regularly update its security protocols to mitigate emerging threats and vulnerabilities.	SCE-01, SCE-02, SCE-03, SCE-04, SCE-05, SCE-06, SCE-07, SCE-08, SCE-09, SCE-10, SCE-11	Security	1- Critical
NFR025	AI-DAPT should perform automated security audits to ensure compliance with industry standards and regulations.	SCE-01, SCE-02, SCE-03, SCE-04, SCE-05, SCE-06, SCE-07, SCE-08, SCE-09, SCE-10, SCE-11	Security	1- Critical
NFR026	AI-DAPT should provide an easy-to-use and user-friendly interface, ensuring a consistent experience across desktop devices.	SCE-01, SCE-02, SCE-03, SCE-04, SCE-05, SCE-06, SCE-07, SCE-08, SCE-09, SCE-10, SCE-11	Usability	4- Low
NFR027	AI-DAPT should feature a responsive design that adapts	SCE-01, SCE-02, SCE-03, SCE-04, SCE-05, SCE-06, SCE-07, SCE-	Usability	4- Low

	to various device screen sizes and resolutions.	08, SCE-09, SCE-10, SCE-11		
NFR028	AI-DAPT should minimize the number of user steps required to perform common tasks through intuitive UI design.	SCE-01, SCE-02, SCE-03, SCE-04, SCE-05, SCE-06, SCE-07, SCE-08, SCE-09, SCE-10, SCE-11	Usability	1- Critical

7 AI-DAPT Draft Minimum Viable Product (MVP)

This section outlines the step-by-step approach taken to identify the list of features and requirements that will be part of the AI-DAPT Minimum Viable Product (MVP). The final set of features and requirements, as identified in this section, will drive the development of the MVP.

7.1 MVP Definition and Approach

The general idea of an MVP is to provide the least developed product that can realize the biggest return on investment with the least amount of risk involved. It facilitates rapid progression to the prototyping phase by focusing on essential features while avoiding those with low user acceptance, high complexity, or misalignment with user needs. While the traditional MVP would simply be one important asset in prototyping, AI-DAPT's MVP plays an expanded role in leading design and development from start to finish. It is also representative of the first release of the platform to be delivered at the end of the project. It reflects the strategy and process for continuous testing, delivering customer value, and validation of ideas and hypotheses regarding the methodology.

The approach to defining the AI-DAPT MVP is provided in Figure 7.1.1 and is divided into three phases: Feature Definition, Feature Assessment, and MVP Consolidation. These phases were conducted in three iterations. It must be underlined that, although an MVP identifies the minimum set of features required for the deployment and validation, it does not mark the end of the development effort. Instead, it acts as a way of releasing high-value components first and allowing more complicated features to be released later, when there is more time for additional research or further refinement of other elements.

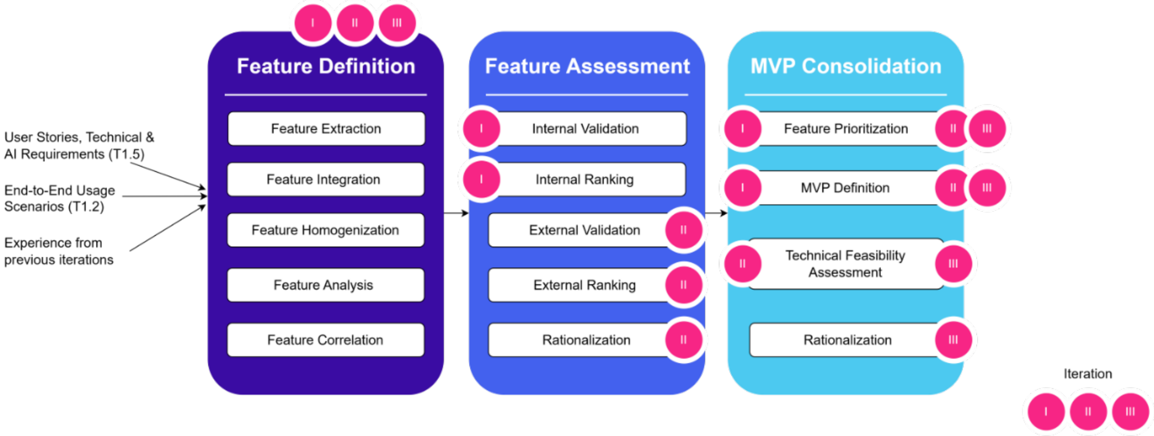


Figure 7.1.1: AI-DAPT MVP Approach

The activities in WP1, WP2, WP4, and WP5 will validate and refine the MVP further with interviews and questionnaires. Another important source of feedback on MVP refinement comes from demonstrator partners, technical partners, and external stakeholders. The iterative approach will be instrumental to mature the platform toward full alignment with the concept and methodology behind AI-DAPT and reaching an effective final solution.

The AI-DAPT MVP is a living document that is constantly fed by new insights and feature requests that create added value. It starts the process with the first iteration aimed at producing the initial draft AI-DAPT MVP, which will be elaborated in the sections below. The main aim is to retrieve a first set of compact features and prioritize them according to their business and technical values. These two values are directly linked to the user stories and end-to-end scenarios mapped to each technical requirement from the previous section. Each MVP requirement consists of one or more technical requirements. The overall priority of an MVP requirement is determined by adopting the highest priority assigned to its associated technical requirements.

7.2 Initial Feature Definition

Based on the user stories developed in D1.1 and the high-level end-to-end scenarios that outline the anticipated use of the AI-DAPT platform from a platform-centric perspective, the relevant technical requirements outlined in Section 6 have been gathered, categorized, prioritized, and structured into the following lists.

Each MVP requirement is assigned a unique ID and title that summarizes its functionality, along with a short description. Technical requirements that are closely related or similar have been consolidated into a single MVP requirement, with the relevant mappings included. The priority of each requirement is determined by adopting the highest priority assigned to its associated technical requirements, as defined in the previous section. Additionally, prerequisites for each requirement have been established, indicating the development of other requirements that must be completed beforehand.

The resulting 50 features are presented below:

ID	PLATF_FR001
Title	Logging Mechanisms
Description	AI-DAPT will provide appropriate logging mechanisms for all operations/components.
Related Technical Requirements	FR001
Priority	1- Critical
Prerequisites	-

ID	PLATF_FR002
Title	Selection and Application of Algorithms
Description	AI-DAPT will provide users with the option to select and apply algorithms from a predefined list.
Related Technical Requirements	FR002
Priority	1- Critical
Prerequisites	PLATF_FR004

ID	PLATF_FR003
Title	Scheduling Operations

Description	AI-DAPT will allow users to schedule operations for automated execution.
Related Technical Requirements	FR003
Priority	1- Critical
Prerequisites	-

ID	PLATF_FR004
Title	Jubyter Notebook Experimentation and Storage
Description	AI-DAPT will allow users to create and experiment with algorithms through the use of Jupyter notebooks, and additionally, it will provide functionality for users to persist/store their experimentations in the form of Jupyter notebook.
Related Technical Requirements	FR004, FR063
Priority	2- High
Prerequisites	PLATF_FR001

ID	PLATF_FR005
Title	Algorithms Management
Description	AI-DAPT will enable users to upload, download, save, and reuse algorithms in their AI pipelines.
Related Technical Requirements	FR005, FR006
Priority	3- Medium
Prerequisites	PLATF_FR004

ID	PLATF_FR006
Title	Integration with Software Platforms
Description	AI-DAPT will support integration with existing software platforms relevant to the users' pipelines.
Related Technical Requirements	FR007
Priority	3- Medium
Prerequisites	PLATF_FR007

ID	PLATF_FR007
Title	API Connections
Description	AI-DAPT will support connection to various APIs for data import from other platforms, devices and services.
Related Technical Requirements	FR008
Priority	1- Critical
Prerequisites	-

ID	PLATF_FR008
Title	Real-time Monitoring
Description	AI-DAPT will allow the users to monitor the progress of their AI pipelines in real-time.
Related Technical Requirements	FR009
Priority	1- Critical
Prerequisites	PLATF_FR001

ID	PLATF_FR009
Title	Notifications
Description	AI-DAPT will provide users with real-time notifications to keep them informed about dataset status updates (including changes to dataset availability, modifications to dataset content, processing status, and errors during operations), AI/Data pipeline stages, and optimizations applied during pipeline execution, such as the use of Adaptive AI techniques.
Related Technical Requirements	FR010, FR011
Priority	1- Critical
Prerequisites	PLATF_FR008

ID	PLATF_FR010
Title	Customizable Data Visualization
Description	AI-DAPT will offer customizable data visualization tools (e.g. dashboards, graphs) to

	help users explore data insights and AI model decisions.
Related Technical Requirements	FR012
Priority	1- Critical
Prerequisites	PLATF_FR012

ID	PLATF_FR011
Title	Explainable AI and Optimization Suggestions
Description	AI-DAPT will support explainable AI (XAI) by clearly explaining automated pipeline steps to users and suggesting data operations to optimize their models and pipelines.
Related Technical Requirements	FR013, FR014
Priority	1- Critical
Prerequisites	PLATF_FR012

ID	PLATF_FR012
Title	Comparison of AI Models
Description	AI-DAPT will provide the ability to compare the results and/or performance of different AI models from previous trials.
Related Technical Requirements	FR015
Priority	3- Medium
Prerequisites	-

ID	PLATF_FR013
Title	Metadata Management
Description	AI-DAPT will provide capabilities to add additional semantic information (metadata) to datasets.
Related Technical Requirements	FR016
Priority	3- Medium
Prerequisites	-

ID	PLATF_FR014
Title	Dataset Recommendations
Description	AI-DAPT will allow datasets and metadata to be catalogued, providing information for evaluating the fitness of the data for its intended uses.
Related Technical Requirements	FR017
Priority	4- Low
Prerequisites	PLATF_FR013

ID	PLATF_FR015
Title	Large Dataset Handling
Description	AI-DAPT will handle and be able to store large datasets.
Related Technical Requirements	FR018
Priority	1- Critical
Prerequisites	-

ID	PLATF_FR016
Title	Workflow and AI Pipeline Templates
Description	AI-DAPT will provide tools for creating, managing, and utilizing shortcuts and workflows, including pre-built templates for AI pipelines and configurable tools for training and inference parameters.
Related Technical Requirements	FR019, FR020, FR021
Priority	2- High
Prerequisites	PLATF_FR020

ID	PLATF_FR019
Title	Dynamic Scaling of Computational Resources
Description	AI-DAPT will dynamically scale computational resources based on pipeline demands to optimize infrastructure usage.
Related Technical Requirements	FR024
Priority	4- Low

Prerequisites	PLATF_FR018

ID	PLATF_FR020
Title	Real-time Training & Inference
Description	AI-DAPT will support real-time training & inference of AI models.
Related Technical Requirements	FR025
Priority	1- Critical
Prerequisites	-

ID	PLATF_FR021
Title	Fundamental AI Methods
Description	AI-DAPT will support fundamental AI methods and techniques, including data annotation, synthetic data generation and bias detection.
Related Technical Requirements	FR026
Priority	1- Critical
Prerequisites	PLATF_FR026, PLATF_FR024

ID	PLATF_FR022
Title	Data Anonymization
Description	AI-DAPT will allow the user perform data anonymization over their datasets.
Related Technical Requirements	FR027
Priority	1- Critical
Prerequisites	PLATF_FR027

ID	PLATF_FR023
Title	Automatic and Manual Data Annotation
Description	AI-DAPT will support both automatic and manual data annotation.
Related Technical Requirements	FR028
Priority	3- Medium

Prerequisites	PLATF_FR024

ID	PLATF_FR024
Title	Full Synthetic Data Generation
Description	AI-DAPT will support synthetic data generation capabilities, enabling full synthetic data generation from scratch and from existing datasets.
Related Technical Requirements	FR029, FR030
Priority	3- Medium
Prerequisites	-

ID	PLATF_FR025
Title	Multiple Data Sampling Methods
Description	AI-DAPT will support multiple data sampling methods (e.g. random sampling, specific interval sampling).
Related Technical Requirements	FR031
Priority	3- Medium
Prerequisites	PLATF_FR026

ID	PLATF_FR026
Title	Data Preparation Tools
Description	AI-DAPT will offer feature engineering and selection tools, including capabilities for missing data imputation, categorical encoding, variable transformation, discretization, outlier engineering, and date/time feature engineering. Additionally, it will provide feature selection capabilities using filter, wrapper and embedded methods to optimize the feature set for machine learning models. Users should be able to configure or select the methods used for each task.
Related Technical Requirements	FR032, FR033
Priority	2- High
Prerequisites	-

ID	PLATF_FR027
Title	Data Cleaning Mechanisms
Description	AI-DAPT will provide data cleaning mechanisms that include handling missing values, removing duplicates, correcting data inconsistencies, and normalizing data formats, with flexibility for user configuration based on data characteristics.
Related Technical Requirements	FR034
Priority	2- High
Prerequisites	-

ID	PLATF_FR028
Title	Human-in-the-Loop and Machine Learning Techniques
Description	AI-DAPT will follow a human-in-the-loop approach for critical data-related tasks such as data cleaning, synthetic data generation, and data annotation, ensuring that users can review, approve, or modify changes before they are applied. This approach supports the use of machine learning techniques and allows the execution of AI pipelines using one or more datasets as input.
Related Technical Requirements	FR035, FR036, FR037
Priority	1- Critical
Prerequisites	PLATF_FR024, PLATF_FR026

ID	PLATF_FR029
Title	File Management Services
Description	AI-DAPT will support file upload, download and conversion services for common text formats (e.g. CSV, JSON, TXT).
Related Technical Requirements	FR038
Priority	1- Critical
Prerequisites	-

ID	PLATF_FR030
Title	Dataset Modification
Description	AI-DAPT will allow users to create new versions of their datasets.
Related Technical Requirements	FR039
Priority	1- Critical
Prerequisites	PLATF_FR027

ID	PLATF_FR031
Title	Data Tagging and Categorization
Description	AI-DAPT will support tags (e.g. real-time data, historical, synthetic, owned) & categories (e.g. operational, medicine) for datasets.
Related Technical Requirements	FR040
Priority	2- High
Prerequisites	PLATF_FR013

ID	PLATF_FR032
Title	Support for Time-Series Data
Description	AI-DAPT will support both time-series and non-time series structured data.
Related Technical Requirements	FR041
Priority	1- Critical
Prerequisites	-

ID	PLATF_FR033
Title	Provision of Baseline Open-Datasets
Description	AI-DAPT will provide baseline open-datasets that users can utilize as examples for their pipelines.
Related Technical Requirements	FR042
Priority	4- Low
Prerequisites	-

ID	PLATF_FR034
Title	Data Confidentiality Levels
Description	AI-DAPT will apply multiple levels of data confidentiality to control access to datasets by the users. Each user must be assigned to a different role, which will determine their ability to upload, view, modify and export datasets.
Related Technical Requirements	FR043
Priority	1- Critical
Prerequisites	PLATF_FR001

ID	PLATF_FR035
Title	Search Functionality
Description	AI-DAPT will support search functionality for datasets, algorithms, and semantic metadata, allowing users to find relevant assets based on their name, category, tag, keyword, date, time, and other semantic information.
Related Technical Requirements	FR044, FR060
Priority	3- Medium
Prerequisites	PLATF_FR013, PLATF_FR034

ID	PLATF_FR036
Title	Combining Multiple Datasets
Description	AI-DAPT will be able to combine multiple datasets into a single, newly created dataset.
Related Technical Requirements	FR045
Priority	2- High
Prerequisites	PLATF_FR030

ID	PLATF_FR037
Title	History Log for Dataset Actions
Description:	AI-DAPT will maintain a history log for dataset actions, tracking events such as when the dataset was created, updated or modified.
Related Technical Requirements	FR046

Priority	3- Medium
Prerequisites	PLATF_FR001

ID	PLATF_FR038
Title	Customizable Notification Preferences
Description	AI-DAPT will allow users to customize their notification preferences, such as selecting the types of notifications they wish to receive, the frequency of notifications and their preferred delivery method (e.g. through the platform, email or both).
Related Technical Requirements	FR047
Priority	1- Critical
Prerequisites	PLATF_FR009

ID	PLATF_FR039
Title	Integration of Hybrid Models
Description	AI-DAPT will allow integration of hybrid models by combining ML & DL algorithms with science-based (first-principles) models.
Related Technical Requirements	FR048
Priority	1- Critical
Prerequisites	PLATF_FR041

ID	PLATF_FR040
Title	Hyperparameter Tuning
Description	AI-DAPT will enable hyperparameter tuning within all models definitions (ML, DL & hybrid).
Related Technical Requirements	FR049
Priority	2- High
Prerequisites	-

ID	PLATF_FR041
Title	AI Model Evaluation and Adaptation Techniques
Description	AI-DAPT will enable the comparison of AI model results and performance across different trials and support adaptive AI techniques for evaluating and optimizing both pre-defined (e.g. holdout, cross-validation, accuracy, etc) and user-defined metrics.
Related Technical Requirements	FR050, FR051
Priority	2- High
Prerequisites	PLATF_FR040

ID	PLATF_FR042
Title	Version Control and Management for AI Models
Description	AI-DAPT will offer version control and management for AI models, facilitating rollback, comparison, and deployment of different model versions.
Related Technical Requirements	FR052
Priority	3- Medium
Prerequisites	PLATF_FR041

ID	PLATF_FR043
Title	Data Upload via Batch Files
Description	AI-DAPT will support data uploads via batch files and datasets.
Related Technical Requirements	FR053
Priority	2- High
Prerequisites	-

ID	PLATF_FR044
Title	Data Import via Streaming
Description	AI-DAPT will support the import of data through streaming mechanisms.
Related Technical Requirements	FR054

Priority	3- Medium
Prerequisites	-

ID	PLATF_FR045
Title	Data Format Validation
Description	AI-DAPT will validate that incoming data is in the correct format.
Related Technical Requirements	FR055
Priority	2- High
Prerequisites	PLATF_FR043, PLATF_FR044

ID	PLATF_FR046
Title	Data Monitoring at Pipeline Stages
Description	AI-DAPT will monitor data at every stage of the AI pipeline to identify potential data issues.
Related Technical Requirements	FR056
Priority	1- Critical
Prerequisites	PLATF_FR001

ID	PLATF_FR047
Title	Model Evaluation Metrics
Description	AI-DAPT will allow users to select and define the metrics used for evaluating their models.
Related Technical Requirements	FR057
Priority	2- High
Prerequisites	PLATF_FR001

ID	PLATF_FR048
Title	Export Data and Pipeline Configurations
Description	AI-DAPT will allow users to export their data and AI pipeline configurations for import into external AI-DAPT compliant AI solutions.
Related Technical Requirements	FR058

Priority	3- Medium
Prerequisites	PLATF_FR001

ID	PLATF_FR049
Title	Model Performance Notifications
Description	AI-DAPT will notify users when their models fail to meet performance thresholds or exhibit suboptimal performance.
Related Technical Requirements	FR059
Priority	2- High
Prerequisites	PLATF_FR001

ID	PLATF_FR050
Title	Partial Synthetic Data Generation
Description	AI-DAPT will support synthetic data generation capabilities, enabling partial synthetic data generation from scratch and from existing datasets.
Related Technical Requirements	FR065, FR030
Priority	3- Medium
Prerequisites	-

ID	PLATF_FR051
Title	Automatic Data Mapping
Description	AI-DAPT will automatically map raw data formats to standard data models using semantic concepts and entities.
Related Technical Requirements	FR061
Priority	2- High
Prerequisites	PLATF_FR035

ID	PLATF_FR052
Title	Dataset Exploration and Visualization

Description	AI-DAPT will allow users to explore and visualize datasets and algorithms through the use of Jupyter notebooks.
Related Technical Requirements	FR062, FR064
Priority	2- High
Prerequisites	-

7.3 Initial Assessment

To validate and prioritize the feature list extracted and defined in Section 7.2, we utilized the MoSCoW methodology [126] in collaboration with the consortium partners. The MoSCoW approach categorizes requirements into four distinct priorities: must-have, should-have, could-have, and won't-have (or will not have at this stage). The validation process ensured that the requirements were not only well-defined but also aligned with the strategic goals of the AI-DAPT platform. This process was carried out through meetings, interviews, and live document reviews, ensuring comprehensive feedback from all partners. Discussions with technical partners were particularly instrumental in aligning the priorities with the platform's requirements and existing rankings derived from the needs of the demonstrators, as already shown in figures 6.1.1 and 6.1.2. This collaborative process enabled the construction of a well-founded prioritization where the "must-have" features represented critical capabilities essential for the MVP, while "should-have" and "could-have" features provided flexibility for future iterations. This process ensured alignment with the strategic objectives of the AI-DAPT platform and provided a structured pathway for development

The assessment of the requirements is presented in the list below:

ID	Title	Assessment
PLATF_FR001	Logging Mechanisms	must-have
PLATF_FR002	Selection and Application of Algorithms	should-have
PLATF_FR003	Scheduling Operations	could-have
PLATF_FR004	Jupyter Notebook Experimentation and Storage	should-have
PLATF_FR005	Algorithms Management	could-have
PLATF_FR006	Integration with Software Platforms	must-have
PLATF_FR007	API Connections	must-have
PLATF_FR008	Real-time Monitoring	should-have
PLATF_FR009	Notifications	could-have
PLATF_FR010	Customizable Data Visualization	must-have
PLATF_FR011	Explainable AI and Optimization Suggestions	must-have
PLATF_FR012	Comparison of AI Models	should-have
PLATF_FR013	Metadata Management	must-have
PLATF_FR014	Dataset Recommendations	could-have
PLATF_FR015	Large Dataset Handling	must-have
PLATF_FR016	Workflow and AI Pipeline Templates	should-have
PLATF_FR019	Dynamic Scaling of Computational Resources	could-have
PLATF_FR020	Real-time Training & Inference	should-have
PLATF_FR021	Fundamental AI Methods	must-have
PLATF_FR022	Data Anonymization	could-have
PLATF_FR023	Automatic and Manual Data Annotation	should-have

PLATF_FR024	Full Synthetic Data Generation	must-have
PLATF_FR025	Multiple Data Sampling Methods	should-have
PLATF_FR026	Data Preparation Tools	must-have
PLATF_FR027	Data Cleaning Mechanisms	must-have
PLATF_FR028	Human-in-the-Loop and Machine Learning Techniques	must-have
PLATF_FR029	File Management Services	should-have
PLATF_FR030	Dataset Modification	should-have
PLATF_FR031	Data Tagging and Categorization	could-have
PLATF_FR032	Support for Time-Series Data	must-have
PLATF_FR033	Provision of Baseline Open-Datasets	could-have
PLATF_FR034	Data Confidentiality Levels	must-have
PLATF_FR035	Search Functionality	could-have
PLATF_FR036	Combining Multiple Datasets	could-have
PLATF_FR037	History Log for Dataset Actions	could-have
PLATF_FR038	Customizable Notification Preferences	could-have
PLATF_FR039	Integration of Hybrid Models	must-have
PLATF_FR040	Hyperparameter Tuning	should-have
PLATF_FR041	AI Model Evaluation and Adaptation Techniques	should-have
PLATF_FR042	Version Control and Management for AI Models	could-have
PLATF_FR043	Data Upload via Batch Files	must-have
PLATF_FR044	Data Import via Streaming	could-have
PLATF_FR045	Data Format Validation	must-have
PLATF_FR046	Data Monitoring at Pipeline Stages	could-have
PLATF_FR047	Model Evaluation Metrics	could-have
PLATF_FR048	Export Data and Pipeline Configurations	could-have
PLATF_FR049	Model Performance Notifications	could-have
PLATF_FR050	Partial Synthetic Data Generation	could-have
PLATF_FR051	Automatic Data Mapping	could-have
PLATF_FR052	Dataset Exploration and Visualization	must-have

7.4 Preliminary AI-DAPT MVP

Based on the preliminary assessment presented in Section 7.3, the Draft AI-DAPT MVP until now consists of a set of features classified as "must-have". In the table below, we display all of them, with the priority being identified from figures 6.1.2 and 6.1.2. In the cases where a feature of the MVP incorporates more than one requirement, the priority is determined by adopting the highest priority assigned to its associated requirements.

ID	Title	Priority
PLATF_FR001	Logging Mechanisms	1- Critical
PLATF_FR006	Integration with Software Platforms	3- Medium
PLATF_FR007	API Connections	1- Critical
PLATF_FR010	Customizable Data Visualization	1- Critical
PLATF_FR011	Explainable AI and Optimization Suggestions	1- Critical

PLATF_FR013	Metadata Management	3- Medium
PLATF_FR015	Large Dataset Handling	1- Critical
PLATF_FR021	Fundamental AI Methods	1- Critical
PLATF_FR024	Full Synthetic Data Generation	3- Medium
PLATF_FR026	Data Preparation Tools	2- High
PLATF_FR027	Data Cleaning Mechanisms	2- High
PLATF_FR028	Human-in-the-Loop and Machine Learning Techniques	1- Critical
PLATF_FR032	Support for Time-Series Data	1- Critical
PLATF_FR034	Data Confidentiality Levels	1- Critical
PLATF_FR039	Integration of Hybrid Models	1- Critical
PLATF_FR043	Data Upload via Batch Files	2- High
PLATF_FR045	Data Format Validation	2- High
PLATF_FR052	Dataset Exploration and Visualization	2- High

Incorporating these essential features into the early release will establish a strong foundation for the AI-DAPT MVP, enabling the project team to showcase the core advantages of the framework. Furthermore, these requirements ensure that the MVP remains flexible and scalable, facilitating ongoing enhancements driven by demonstrators (end-users) feedback and real-world usage.

8 Ethics Analysis, Constraints & Considerations Towards Trustworthiness

8.1 Ethics-by-Design Approach

AI-DAPT Trustworthy Framework under development relies on the Ethics-by-Design approach, which is strongly consistent with the project's human-centric vision, including its efforts towards ensuring explainability in order to promote end-users' trust.

Such an approach is in line with the indications elaborated by the Independent High-Level Expert Group on AI, in particular, their "Ethics Guidelines for Trustworthy AI" [127] and ALTAI assessment list [128], besides being based on the guidance document elaborated by the EC "Ethics By Design and Ethics of Use Approaches for Artificial Intelligence" [129]. Such guidance document concerns all research activities involving the development or/and use of AI-based systems or techniques and was elaborated to support innovators and developers in adopting an ethically-focused approach while designing, developing, and deploying and/or using AI based solutions.

The AI-DAPT Consortium will adhere to the ethical principles outlined in the guidance document for the AI systems' development and use: the design activities in AI-DAPT include and will continue to include the appropriate considerations of the key characteristics that an AI-based system/application must have in order to preserve such principles and ensure the protection of the fundamental rights, as enshrined in the Charter of Fundamental Rights of the European Union (EU Charter). Therefore, the AI-DAPT Consortium will continue to consider the following ethical principles, embedding into the AI system/s under development the corresponding general ethics requirements (see section 8.3 of these document):

I. Respect for Human Agency

According to this principle, the human beings must be respected to allow them to make their own decisions and carry out their own actions. It comprises the following specific principles, defining fundamental human rights:

- Respect for human autonomy, to let the individual think, decide and chose for himself/herself. The AI system might restrict the human autonomy without any malevolent intent, but simply as a consequence of a lack of understanding about peoples' values and preferences.
- Respect for dignity, to not compromise the intrinsic worth that every human being possesses. People should be always treated with respect, including when using or being subjected to AI-based systems, without instrumentalizing, objectifying or dehumanizing them;
- Respect for freedom in its several facets, such as the freedom of movement, the freedom of speech, the freedom of access to information, and the freedom of assembly. Any form of coercion, deception, exploitation of vulnerabilities, and manipulation, should be avoided.

II. Privacy and Data governance

In order to respect the right to privacy and data protection, the AI system should be built following the data minimization principle and the privacy and data protection by design and by default, which are also set forth by the General Data Protection Regulation (GDPR). The surrounding data governance models, also in view of fostering trust in data sharing, should ensure the privacy protection, the data accuracy and representativeness, besides protecting personal data and enabling humans to actively manage their own data and the way the system uses them. Some ethical issues could arise also in case the AI system uses non-personal data. It is also key to consider the differences and between the different types of data and data models used in AI systems, since each category raises different challenges: training data, model data, production data, knowledge data or analysis-to-data, data-to-analysis and data-and-analysis-to-lake models.

III. Fairness

This principle asks that people are given equal rights and opportunities and are not advantaged or disadvantaged undeservedly. It is necessary to avoid any forms of discrimination on the basis of the fundamental aspects of one's own identity which are inalienable and cannot be taken away, such as gender, race, age, sexual orientation, national origin, religion, health and disability but even people's diverse personalities, experiences, cultural backgrounds, cognitive styles, and other variables that influence personal perspectives. The diversity should be supported, accommodating for these differences. The procedural dimension of fairness requires that the procedure is not designed in a way that disadvantages single individuals or groups specifically. The substantive dimension of fairness requires that the AI system does not foster discrimination patterns that unduly burden individuals and/or groups for their specific vulnerability.

IV. Individual, Social and Environmental Well-being

Any AI system should contribute to, and not harm:

- individual well-being, so that people can live fulfilling lives, pursuing their own needs and desires in mutual respect.
- social well-being, concerning the flourishing of societies, whose basic institutions function well and where sources of social conflict are minimized.
- environmental wellbeing, which regards the well-functioning of ecosystems, the sustainability, and the minimization of environmental degradation.

The end-users affected individuals and communities and relevant stakeholders should be identified at the very beginning of the AI system design, in order to make possible a realistic assessment of how the AI system could enhance or harm their well-being.

V. Transparency, Explainability and Objection

The purpose, inputs, and operations of AI programs should be knowable and understandable to its stakeholders, including the data, the system, and the processes by which it is designed and operated. It is key that the stakeholders are able to understand the main concepts behind it, including how, and for what purpose, the system functions and comes to its decisions. On this regard, a potential issue is the possible trade-off between the transparency and the IP rights / confidentiality / trade secrets: to tackle with it, the selective transparency (e.g. confidentially to trustworthy third parties) and technology or confidentiality commitments should be considered. The transparency principle is strictly interrelated to the respect for human agency.

VI. Accountability-by-Design, Control and Oversight

The human being must be able to understand, supervise and control the design and operation of AI based systems and the actors involved in their development or operation must take responsibility for the way that these applications function and for the resulting consequences. The accountability principle is interrelated with the transparency and oversight: in fact, on the one hand, in order to be held to account, developers or operators of the AI system must be able to explain how and why a system has a particular feature or results in certain outcomes and, on the other hand, the effective human oversight requires that the human actors are able to understand, supervise and control the design and operation of the AI system.

The adherence to the Ethics by Design approach will allow the AI-DAPT technical/research team to think about and address potential ethical concerns at the initial stage of the design and development of the project's system, embedding the ethical principles and corresponding requirements into the design process by considering them as system requirements, together with the other technical, functional, and non-functional requirements.

In AI-DAPT, the use of the Ethics by Design approach will be coupled with the Ethics and Data Protection Impact Assessment Methodology in WP5: this methodology includes a component specifically dedicated to the AI systems. In this way, it will be ensured that all major AI ethics principles are adequately considered during the design and development of the AI-DAPT AI artefacts, together with the attention to legal compliance. More details on the Ethics and Data Protection Impact Assessment Methodology are provided in Section 8.4 of this deliverable.

Thanks to the use of the Ethics by Design Approach, the AI-DAPT technical team will be supported by the legal and ethical requirements in thinking about ethics and legal compliance, whilst conceiving and developing the AI artefacts: in this way, it will be clear for them what is required, which are the legal and ethical boundaries of the activities and what specific work must be undertaken to ensure the legitimacy and fairness of AI-DAPT solutions and pilots.

8.2 Ethical principles, regulatory landscape and ethical and legal requirements for AI-DAPT Framework

8.2.1 Key aspects and challenges

The project embraces a data-centric approach in AI, combined with hybrid science-guided ML approaches, in order to introduce automation and AI-based systematic methods to support the design, the execution, the observability and the lifecycle management phases of data-AI Pipelines, capable of continuously learning and adapting on the basis on their context. AI-DAPT leverages automation and AI techniques towards constructing robust, intelligent and scalable data-AI pipelines, which can adapt and learn from their environment and execute efficient steps that integrate operational and business logic. The data-centric approach in AI rotate around robust, intelligent and scalable data-AI pipelines, containing a set of efficient data-driven (besides model driven) steps for introducing automated, intelligent end-to-end "Data for AI" Pipelines, which will enable the proper purposing, collection, documentation, (bias) valuation, annotation, curation and synthetic generation of data to reach to the right data for the right problem, with a fusion of science-guided and AI models.

A core data pipeline typically includes five phases, which can collectively appear in a combined data/AI pipeline and are supported through manual (M) or automated (A), typically AI-based, methods [130]:

- Phase I - "**Data Design for AI**": in this phase the data scientists select suitable data for the AI solution, drawing on domain knowledge. Automated processes fetch raw data from internal databases and the data characteristics are analysed and summarized collaboratively by data scientists and business users and findings are produced.
- Phase II - "**Data Sculpting/Nurturing/Curation for AI**": in this phase the data representativeness and quality is ensured by using AI/ML techniques. Features are annotated semantically and engineered, with relevant ones chosen for the AI model. Data quality is enhanced thorough cleaning techniques.

- Phase III - "**Data Generation for AI**": this phase is devoted to create synthetic data to supplement or replace real data. Data utility assessment evaluates the suitability of synthetic data.
- Phase IV - "**Model Delivery for AI**": during this step, the data scientists oversee the AI model lifecycle, using hybrid science-guided ML approaches. Models are configured, trained, and deployed for real-world application, considering prediction uncertainty.
- Phase V - "**Data-Model Optimization for AI**": this phase concerns the continuous monitoring and improvement of the AI solution based on real-life operation circumstances. Data and model observability ensure timely adjustments.

Therefore, the key legal and ethical implications (and, in some cases, challenges) pertaining to the AI-DAPT system regard data and AI. In AI DAPT, the approach chosen to address these implications is driven by the cross-fertilization of, on the one hand, technology and, on the other hand, law and ethics. The attention to the legal and ethical implications is in line with the Consortium's aim to prioritize at the maximum possible extent the human well-being when teaming with machines and AI-based solutions, putting the innovations at the service of human needs and interests, fostering his/her flourishing, empowerment, enhancement and augmentation, so to give rise to really inclusive solutions. This is at the centre of the AI-DAPT Trustworthy framework, in coherence with the human-centric approach fostered by the European Commission.

The key legal and ethical implications of AI-DAPT technology are described below.

<u>Privacy, Data Protection and "Data Ownership"</u>

The privacy implications relevant to AI-DAPT, and especially to the human-machine-interaction technologies and human-centric aspects of its AI-based system, broadly refer to the possibility to withhold personal information and demand for its use with consent. In a human-machine collaboration environment it is paramount to ensure i) that the operator is aware that his/her data are being (or might be) collected, ii) that he/she gives informed consent to this, iii) that data are stored in a secure manner, iv) and that the ethical requirements for data collection, storage and use are met.

Depending on the context, individuals expect different levels of privacy, as shown, for instance, by the AI-DAPT demonstrators.

In AI DAPT and in line with the concept of trustworthy AI paradigm and the HITL and XAI approaches, the full exploitation of the potentialities of the new scientific standpoint for continuous optimization of the entire lifecycle management of data-AI pipelines entails the data collection, processing and use in relation, for instance, to the real-time monitoring solution of the human-centred processes, as well as in general in the data collection and treatment steps of AI-DAPT Phases I & II and within the cloud-based AI-Ops Platform. Such data collection, processing and use might also include personal data, in some cases. In fact, it should be considered also that devices and machines are often capable of recording everything, being often equipped with sensors. In this context where machines can capture data about people, besides about equipment and environments, the main challenge lies in using data ethically and in an informed manner. Nevertheless, despite in some cases it might result complex to gather informed consent according to the European Data Protection Framework (GDPR), it is essential that the human beings are clearly aware that data were being collected, handled and stored. As regards the piloting operations, the AI-DAPT Consortium is committed in ensuring this. The collection, processing and sharing of any personal data (e.g. of data subjects) will be analysed under the principles of the GDPR and the proposed Data Governance Act. Secure and privacy-preserving solutions "by design" will be provided, so that data subjects can trustfully and securely channel and share their data with the support AI-DAPT. The de-anonymization/re-identification challenge will be addressed as well, considering also that such challenge is reinforced by the increasing scale, speed and complexity of data analytics in AI-DAPT.

Many current pseudonymisation and anonymisation approaches are insufficient, since the combinations of non-personal data might lead to the identification of persons and/or other sensitive information: in other words, by matching anonymous data (de-identified data) with publicly available information, or auxiliary data, in some cases it is possible to discover the person to whom the data belongs. This point is crucial as personal data, once anonymized (or pseudo-anonymized), may be freely processed, without any prior authorization by the data subject. Considering this, it is key that the AI-DAPT Consortium adopts GDPR-compliant pseudonymization. It is suggested to reduce the risk of re-identification through the use of separately kept "additional information", required for re-identification. Besides conforming to the applicable legislation, data collection and usage should, on the technical side, control the granularity at which data is shared with third parties throughout the data lifecycle, as well as enforce that the data are only used for the defined purpose.

Another aspect relevant to AI DAPT is Data Ownership [131], which represents an inherently problematic concept from a legal perspective, lacking both a legal definition at the EU level (there is even no reference to “data ownership” as such in the law) and a common understanding in legal literature on what the scope of such an ownership right would be. This situation is likely due to the non-exclusive and inexhaustible nature of data. The concept of “Data Ownership” is often used to describe that a certain party has a certain control over specific data, without specifying what this ownership right entails exactly. In fact, the central point for data is not who owns them (if anyone), but who has the right and ability to access and use them. Acknowledging this, the Data Act focuses especially on access and usage rights, without introducing a common concept of data ownership: it considers the data holders and grants certain access and usage rights in relation to them, representing the most recent and largest scale EU level attempt to streamline access and usage rights in a more consistent and homogenous way Likewise, smaller scale and sector specific initiatives are focused on access and usage rights, rather than on data ownership. This approach is also common to personal data: in the GDPR the right to data protection clearly grants data subjects a significant degree of control over their personal data, which are different from a data ownership right (meant as an exclusive right to use, and control the use of, certain data), since such a control is not absolute. Over the past few years the EU legislation regulated data access and usage rights as a nuanced evolution of data ownership concepts, thereby focusing on the right to access and use certain data for certain purposes described by law, without granting or defining ownership rights as such. Usage and access rights have been enshrined, besides in the Data Act and in the GDPR, in other legislations, such as the PSD2 Directive in the banking legislation.

Synthetic data generation, accuracy, representativeness and Generative AI

Synthetic data are artificial data generated from original data and a model that is trained to reproduce the characteristics and structure of the original data or, in other words, computer-generated information for testing and training AI models. The utility of the method and model to generate synthetic data through the generation process (the so-called synthesis), determine the degree to which there are an accurate proxy for the original data. Synthetic data can be generated employing real datasets, knowledge-gathered by the analysts, or a combination of these two.

The generation and use of synthetic data are gaining traction within the machine learning domain, since it helps training machine learning algorithms that need a large volume of labelled training data, which in some cases could not exist or could be expensive or impractical or have data usage restrictions (for instance due to data protection/privacy implications, such as, for instance, in the health-care domain). The real data, i.e. the data collected directly from the real worlds, has inherent limitations and incompleteness, raising issues such as data imbalance and data discrimination in practical applications. Considering these drawbacks and the difficulties to satisfy the demand relying solely on real data, diverse methods are employed for generating synthetic data through existing real data [132].The synthesis is usually relatively cheap and give rise to automatically labelled data, besides address many of the privacy and data protection concerns occurring when training AI models on real-world cases. Thanks to this process, larger datasets that would not even be available using real data can be produced. The generation and use of synthetic data have also a positive foreseen impact on privacy and data protection and is aligned with the data protection by design and by default approach, followed by AI-DAPT Consortium, since the personal data of individual do not have to be disclosed. In other words, synthetic data represent a less privacy-intrusive way to train the ML/AI models, due to the fact that the data used in the training process does not directly refers to an identified or identifiable person. They can be considered as a Privacy Enhancing Technology (PET), enhancing the privacy in data transfer [133]. On the other hand, the synthetic data generation and use might improve the fairness of the overall solution, since they might contribute to mitigate bias by using fair synthetic datasets to train the AI models, including using datasets with a better representativeness of the world, for instance avoiding gender-based or racial discrimination. Nevertheless, this kind of manipulation of datasets to create fair synthetic datasets might result in inaccurate data. Furthermore, additional challenges regard the fact that the quality of the model depends on the data source, being the quality of synthetic data strongly correlated with the quality of the original data and the data generation model. This also implies that synthetic data may reflect the biases in original data.   On the other hand, the synthetic data may not cover some outliers that original data have. In other words, due to their flexible nature, the synthetic data might be biased in behaviour or inconsistent or inaccurate in comparison to real datasets. The synthetic data might also pose some challenges regarding data protection, mainly due to the risk of reidentification, especially in cases where, to preserve the maximum utility of the data, the synthetic dataset mimic at the maximum extent the real data, thereby revealing about real people, with risks to privacy and other human rights. There is also the risk of membership inference attacks, i.e. the possibility for an attacker to infer whether the data sample is in the target classifier training dataset). This applies especially in case of outlier records, i.e., data with characteristics that stand out among other records [133]. The generation and use of synthetic datasets in AI pose a range of potential risks and challenges that require careful assessment and management. In fact, the synthetic data generation process often harbours issues, especially pertaining to the fairness and representativeness of data distribution, which might affect the performance of models and potentially lead to biases and discriminatory practices in real-world applications: the choices made in dataset compilation have a profound impact on the AI's behaviour, echoing the individual' conscious and unconscious biases. It is paramount integrate critical filtering, ethical considerations, cultural sensitivity, and diversity into every stage of AI development, starting from dataset creation [132]. The synthesis requires, besides a deep understanding of statistical methodologies, the appropriate consideration of the ethical and legal implications, including an active effort to identify and mitigate any potential biases, in order to ensure fairness and representativeness in the datasets. The main challenges, shortcomings or risks in creating and using synthetic data are as follows:

- Insufficient representativeness of the synthetic datasets: the generation of synthetic data often doesn't adequately consider demographic diversity. This might lead to unbalanced data distributions in terms of gender, age, race, etc. For example, if, during the creation of the dataset, the data primarily regard individuals from specific racial or age groups, the trained models might perform poorly when dealing with other groups, potentially with severe unfair or discriminatory outcomes in real-world applications, such as certain age groups being incorrectly identified or entirely overlooked. The biases in the data can give rise to overrepresentation or underrepresentation of specific groups in the resulting synthetic datasets, leading to skewed perceptions and decisions made by AI systems, which, in turn, might inadvertently become a medium through which the biases are perpetuated and amplified. Issues might occur particularly regarding how data generation algorithms process and interpret the input data: in this process, such algorithms might inadvertently learn and replicate prejudices, stereotypes and biases (such as gender or racial biases) that are inherent in real-world data sources and, in this way, the biases and prejudices can infiltrate the process of AI's decision-making. The pertinent statistical attributes of synthetic datasets, compared to their authentic counterparts, should be preserved, in order to avoid the potential proclivity for models to engender misleading prognostications within practical applications, which compromise their fidelity to encapsulate real-world phenomena.

- Incomplete and/or inaccurate data: as regards incompleteness, the lacunae or partial information within synthetic datasets, resulting from imperfections or errors in capturing the manifold changes inherent in authentic datasets during the synthetic generation process might hinder the model's capacity to accurately prognosticate or manage scenarios and influence its resilience and pragmatic utility. As regards inaccuracy, due to the errors or inaccuracies within the synthetic datasets, compared to the veracity of real-world datasets, caused by algorithmic imperfections or other contributory factors, erroneous patterns might be internalized, thereby inducing biased predictions and undermining the overall performance and reliability of the model when confronted with genuine data.

- Undue sterility and inconsistency: even the paucity of features like multifarious noise and intricacies inherent in authentic data (where, in authentic scenarios, data encapsulate diverse interferences, errors and uncertainties), might lead to an undue sterility, hampering the model's efficacy within realistic environments. Synthetic datasets often fail to encapsulate the rich tapestry inherent in authentic datasets, which often embody variations stemming from diverse sources, temporal epochs, and environmental conditions. This result in paucity of inconsistency within synthetic datasets, compared to authentic datasets, which might lead to engender challenges for models in adapting to the multifaceted vicissitudes originating from disparate sources, environmental conditions, temporal epochs, thereby decreasing the generalization performance vis-`a-vis diverse datasets.

- Over-smoothing and/or neglecting temporal and dynamic aspects: certain model for synthetic data generation, may overly attenuate or oversimplify the data, which produce an attenuated representation devoid of the nuanced details and diversity inherent in real-world datasets, thereby introducing challenges for the model in assimilating complex variations within genuine data. On the other hand, certain model might inadequately capture facets inherently pivotal within authentic datasets, such as temporal and dynamic nuances, thereby resulting in ineffectuality of models in real-world applications.

- Liability for AI-generated data: this regards the AI responsibility in generating fictional content. Some authors argue that there might be liability challenges associated to the generation of synthetic data and content, potentially leading to misinformation, misunderstandings, or the dissemination of false information with detrimental societal impacts. This is related to the use of Generative AI techniques and the debate regarding the responsibility for AI-generated content. The question regards who is responsible in the case of wrong output of the AI system, due to the involvement of multiple partiers in developing and deploying the AI system, as well as the user's influence on the output by inputting the prompt.

- Security and Adversarial Attack Risks: these risks regard the possible malicious use of synthetic data, which might render AI models unstable during adversarial attacks, since the surrounding models may not adequately learn the complexity and diversity of the real-world during training. This increases the likelihood of deception or manipulation, and thereby pose threats to the credibility and security of the AI system.

Risk of technological determinism [134]

There is the pressing need to achieve social control over AI, which represent a technology appearing to be as autonomous as no other. The efforts at the social control of technology are nothing new. Nevertheless, Artificial Intelligence, with its unique nature, appears the most resistant to such control: this validates the amount of attention the question it receives in view of avoiding that the future society is determined by the nature of this technology. The attitude/historiographic methodology named "technological determinism", which was widely criticized and almost completely disregarded since the second half of the 20th century, is recurrent again in the case of AI, especially in relation to the emergence of Generative AI solutions. The debate on technological determinism is a historical and ongoing point in social and technology research [135]. Despite there isn't a canon definition of technological determinism, it is common understanding that it refers to the notion that technology shapes society and culture, as well as shapes and control human behaviour, rather than the other way around (i.e. simply being a tool that we use to achieve human goals). One of the main potential pitfalls of technological determinism is that this position could lead to a scenario where technology is seen as an unstoppable force, beyond our control or regulation, as if it operates independently of human influence. It also posits that technology has a deterministic effect on society, driving social change in a specific direction that cannot be altered without fundamentally changing the technology itself. The AI revolution and the ever-evolving landscape surrounding it have brought the principles of technological determinism to the forefront. AI technologies are reshaping our industries, economies, and daily lives and societal norms, embodying the deterministic effect of technology on society. It is therefore core to harness the power of technology for the greater good while mitigating its potential dangers, and to shape the direction of technological development to align with societal values and needs. The ethical mandates, such as the Ethics Guidelines for Trustworthy AI and the Guidelines on the responsible use of generative AI in research, as well as the regulatory developments, such as the AI Act and the AI Liability Directive, are moving in this direction. In particular, it is particularly relevant for AI-DAPT, the Human Agency and Oversight. They are functional to ensure that the AI system supports human autonomy and decision-making, in line with principle of respect for human autonomy which must be central to the system's functionality. This is related to the Human in the Loop Approach adopted in AI-DAPT and the commitment towards Explainable AI in the project.

Ethics concerns associated to the adaptiveness of the AI system

Another aspect to investigate from the legal and ethical viewpoint in relation to AI-DAPT concerns the adaptive nature of AI, which allows the system to modify its behaviour in response to environmental input, so that the system itself can continuously adapt and learn on its own, dynamically adapting to the intricacies of the environment. This allows the improvement of performance through hands-on learning, just like human cognition is adaptive, and swiftly adjusting to a variety of situations. Adaptive AI allows the intelligent systems to dynamically adapt to the intricacies of the environment [136]. Although adaptive AI shows great potential, there are obstacles and constraints as well to its growth and use. The effectiveness of the Adaptive AI solution is impacted by the volume and quality of the data, since incomplete or skewed datasets may hinder flexibility. Adaptive AI systems ought to be able to justify their choices and flexibility (Explainability and Interpretability), but it might be difficult to make sophisticated adaptive models explainable. Research on interpreting highly adaptable systems' decision-making processes is still in progress. Furthermore, adaptive AI systems must conform to moral principles and cultural norms and it might be difficult to ensure an ethical behaviour and bias prevention in systems that are always changing.

A growing body of research [137] [138] regards the development of self-adaptive framework for responsible Adaptive AI system / LLM use: in other words, this trend concerns the development of self-adaptive frameworks to align adaptive AI systems / LLMs behaviour with ethical principles, spanning from dynamic ethical frameworks to autonomous self-improvement for ensuring that the adaptive AI System / LLMs autonomously align their behaviour with ethical principles in a self-adaptive manner. They emphasize the need for dynamic ethical guidelines capable to adapt to context and user feedback and that future LLMs should be able to autonomously dynamically adjust their behaviour to align with ethical values (besides with context and feedback), incorporating autonomous self-improvement guided by ethical principles. The self-adaptive behaviour is expected to be able to ensure that AI systems remain aligned with human values over time. The potential components of a self-adaptive Responsible AI Framework include:

i. **Value alignment and ethical foundation** of the framework itself, grounding it on the principles of fairness, transparency, and accountability. It is necessary to embed human values and ethics into the AI systems, aligning them with human values.

ii. **Continuous monitoring and evaluation** the output of the AI system/LLMs to timely identify and address possible ethical deviations, as well as to assess the AI system/LLMs against evolving ethical standards. Methodologies should be developed on this purpose, involving both automated testing/processes and human oversight/judgement.

iii. **Quantifiable measures of ethical behaviours**, including metrics and feedback loop, capable of guiding the system's self-adaptation, prepared in collaboration with human/ethics experts.

iv. **Human-in-the-Loop** approaches to ensure that the AI system's autonomy is balanced with human control, by allowing the human expert to intervene and fine-tune the adaptation process. In fact, also in case of a self-adaptive Responsible AI Framework, the role of human involvement is critical and the balance between human oversight and AI autonomy is a key aspect in the self-adaptation process.

v. **Explainability** (XAI), to make the AI system/LLMs' decision-making processes more transparent to human users and developers.

Safety, Liability and Accountability

In line with the considerations of the "Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics" [139], it is widely acknowledged that the specific characteristics of the autonomous and adaptive AI systems, such as the potential (partial) opacity of AI decision-making and lack of explainability, the autonomous behaviour, the self-learning capabilities, the continuous adaptation, the data dependency and the limited predictability raise challenges regarding the liability domain, making it difficult to predict the behaviour of an AI-enabled product and to understand the potential causes of a damage, as well as to meet the burden of proof for a successful claim. The level of complexity is increased by the plurality of economic operators involved in the supply chain, by the openness to updates and upgrades after the placement of an AI system on the market, as well as by the multiplicity of components, software tools and services working together within the new AI-empowered technological ecosystems, which, due to the connectivity of their IoT products and openness, can be exposed to additional cyber-threats [139]. Likewise, the EU Parliament [140] stated that "the complexity, connectivity, opacity, vulnerability, the capacity of being modified through updates, the capacity for self-learning and the potential autonomy of AI systems, as well as the multitude of actors involved represent nevertheless a significant challenge to the effectiveness of Union and national liability framework provisions". As underlined by the White Paper on Artificial Intelligence – A European Approach to excellence and Trust [141] and the Report on safety and liability [139], the AI growth and its wide uptake relies also on the adequate instruments and approaches to address the liability aspects at policy level, especially the liability for damage caused by AI-systems, including the uncertainty regarding the allocation of responsibilities between different actors. In the same direction, the Resolutions adopted by the European Parliament in October 2020 [142] on ethics and civil liability related to the AI systems, underlined the need of harmonization of the legal framework for civil liability claims and of a regime of strict liability on operators of high-risk AI systems. The European Parliament, on the other hand, acknowledged the specific and coordinated adjustments to the "liability regimes are necessary to avoid a situation in which persons who suffer harm or whose property is damaged end up without compensation" [142]. As examples of potential harms, the following can be mentioned: unavoidable or inherent harms, defect-driven harms, misuse harms, unforeseen harms, systemic harms, as well as collateral harms.

It is paramount the clear understanding of responsibilities between different actors when using an AI system, like in AI DAPT, in case it makes mistakes producing damages or injury to property and human beings. However, in case of harm, questions of attribution and remedies might arise at the intersection of products liability and AI, such as questions on the attribution for AI-induced harms and the identification adequate mechanisms to mitigate possible AI harms. Typical questions are: "Whose fault is it if an AI system takes a decision which causes harm?", "How to apportion such a fault?", "How to avoid the repetition of such mistakes in the future?" As highlighted, in case of the AI DAPT solutions, the difficulties are increased by the fact that the AI systems are able to adapt and learn, going beyond the simple implementation of human-designed algorithms. Whilst in some cases the harm caused by an AI system could be the direct consequence of its programming or of its negligent design, training, or operation (e.g., lack of adequate cybersecurity protections), in other cases, when the products evolved, it is not easy, in relation to products liability, to understand whether the developers need to bear responsibility for the AI products they create, even when those products evolve in ways not specifically desired or foreseeable by their manufacturers, as well as to apportion blame and responsibilities when there are multiple companies that have had a hand in designing an AI system (or in shaping the post-sale algorithm evolution). It should be also considered that the liability risk associated with AI systems might differ depending on the function of the AI output: for instance, the predictive systems differ from fully autonomous systems, where humans are mainly "out of the loop". The existing liability reference framework includes: i) National liability regimes, which are still fragmented, lacking of clear liability rules specifically applicable to damage resulting from the use of emerging digital technologies such as AI (with the limited exception of highly or fully automated vehicles); : i) the partially harmonised EU legislation on liability for defective products (Directive 85/374/EEC, applicable to a vast range of products, including complex AI-driven devices; and iii) the AI Act [143], which, regarding the civil liability for AI systems, imposes specific obligations upon providers, importers, users, distributors, and even third parties (Articles 16 to 29), following a risk-based approach. In particular, the AI applications are classified according to a typology of risks from none to high-risk, in line with both the Report on the civil liability regime for AI and the White Paper. Furthermore, the evolving regulatory framework under development on the civil liability for AI systems includes:

- AI Liability Directive (AILD) Proposal

  This proposal will complement the AI Act, providing a new set of liability rules specifically targeted at AI, tackling consumers' liability claims for damage caused by AI-enabled products and services. It is directed, on the one hand, i) to make it easier for victims of AI-related damage to get compensation and to ensure that victims benefit from the same standard of protection across the EU when harmed by AI products or services; and, on the other hand, ii) to ensure an easier access to redress for the victims and provides broader protection for victims (individuals or businesses), whilst increasing guarantees.

- Revised Product Liability Directive (RPLD) Proposal [144]

  This proposal, together with the AI Liability Directive, is directed to adapt liability rules to the digital age, circular economy and the impact of global value chains, in particular by modernizing the liability rules for products in the digital age, with a focus on the strict liability of manufacturers for defective products. It updates and reinforce the existing rules, based on the strict liability of manufacturers, for the compensation of personal injury, damage to property or data loss caused by unsafe products. This instrument is expected that it will provide the businesses with legal certainty and will ensure that victims get fair compensation when defective products cause harm. It is applicable to all tangible and intangible unsafe products, including embedded or standalone software and digital services necessary for the products' functioning.

Despite both the RPLD and the AILD reverse the burden of proof for damage caused by AI applications (such as cobots) under certain conditions, nevertheless, they differ because the former is based on strict liability (meaning the presumption of malfunctioning applies under specific condition, and constitutes a legal basis for claims), whilst the latter merely harmonises certain aspects of legal proceedings initiated under national fault-based liability regimes and requires the complaint to prove the defendant is at fault for breaching the requirements of the AI Act.

The liability aspects are strictly interrelated with the **accountability principle**.

In fact, AI accountability refers to the idea that AI systems should be developed, deployed, and utilized in a way that ensures that the responsibility for bad outcomes can be assigned to liable parties.

On the other hand, accountability concerns in relation to the AI-enabled technology often are due to the opacity and complexity of machine and deep learning systems, the number of stakeholders typically involved in the value chain for developing, implementing and using AI products, and the AI dynamic learning potential. In case of "black box" AI systems, where the process behind how an output was achieved can't be fully explained or interpreted by its users, it is difficult both to assign responsibility and to hold parties accountable for harmful outputs.

Accountability asks to set mechanisms to ensure the ability to review AI systems, meaning the ability that AI systems and their outcomes, both before and after their development, deployment and use, can be observed and analysed. Accountability is interrelated with traceability [145], which requires to keep a complete and clear documentation of the data, processes, artefacts and actors involved in the entire lifecycle of an AI model. Pursuant to the AI Act, in case something goes wrong with an AI system, there should be someone responsible. In particular, it introduces the concept of "provider accountability", which means that the individuals or organizations developing, deploying, or operating AI systems are held responsible for their actions. Article 17 "Quality Management System", in relation to the providers of high-risk AI systems, specifies that they "shall put a quality management system in place that ensures compliance with this Regulation. That system shall be documented in a systematic and orderly manner in the form of written policies, procedures and instructions, and shall include" at least the aspects listed in the articles itself.

The synergy between the Human in the loop Approach and Explainability [146] **and the complementary role of the hybrid science guided models**

One of the primary concerns regarding the AI system regards the potential for them to perpetuate biases and discriminatory practices, since they learn from historical data which may contain inherent prejudices. The human-in-the-loop (HITL) approach, incorporating human expertise and values into AI algorithms and involving the integration of human judgment and oversight into the AI systems, has emerged as a potential solution to tackle with the risks of biased or opaque decision-making and related challenges, like this, associated with the AI adoption. The HITL [147] approach consists in a method of having humans involve and interpret the output of AI models or, in other words, a new paradigm in which users are engaged to improve and personalize automatic AI-based solutions. In this operating model the AI decisions and actions are supervised and, if necessary, modified or validated by humans, avoiding a fully automated approach by highlighting the importance of human intervention in AI-based decision-making processes. This approach, connecting humans to the model loop in a specific way, so that the machine can learn human knowledge and experience during the loop, is acknowledged as useful for improving the transparency (in particular the explainability), the accountability and the performance of AI systems, and ultimately for building trust with end-users and foster the adoption of such systems. This approach is helpful for adapting and reconceptualizing the users' role in these systems, considering human factors in designing more effective and flexible collaborative systems between humans and machines [148] and valorising the role of the user in his/her interaction with the learning system, as well as the providing of feedback, guidance, or input when needed. Multiple researchers apply human-in-the-loop based techniques to optimize models from the perspective of data. Others apply human intervention in dialogue to enable machines to learn human intelligence from dialogue procedures iteratively: in other words, they integrate humans into the reasoning loop in order to allow the machines also to learn more about human experience. Such a method brings a number of benefits [149], such as:

- The valorisation of **human intuition in the contextual understanding**, since human beings still retain an irreplaceable advantage in understanding complex and variable contexts, thanks to their ability to assess new situations and interpret cultural or emotional nuances.
- The **human ability to manage exception**, since in unforeseen situations or in cases falling outside standard AI learning patterns, human intervention provides greater flexibility, as well as a more adaptive and personalized response to specific needs.
- The **alignment with the ethical, social and legal values of society,** since the human presence ensures that the decisions made by the AI system, which at the moment doesn't possess a sense of ethics or an understanding of the moral implications of its actions, are consistent with such values. This is also expected to foster trust and a better acceptance of the AI system. More in general, the HITL approach has the potential to address concerns about the fairness, accountability, and transparency of AI systems, helping, thanks to the involvement of the human perspectives in the AI development process, to identify and mitigate biases, thereby ensuring that the AI algorithms are more representative and equitable. The HITL approach is useful to **prevent bias,** being the human oversight crucial for addressing biases in AI algorithms, so that to ensure that decisions align with ethical standards and societal values, promoting fairness in AI applications.
- **Customization/fine tuning of the AI system** by the end-users (e.g., through re-training and debugging of models/classifiers) to suite their individual needs/expectations in different contexts better could improve the perceived utility of the system and its acceptance, the detection capabilities of the AI system for each user, helping to address a wider diversity of user profiles and expectations and, at the same time, increase the flexibility for adapting to most of the system's functionality to the users' needs and overall performance of the system. By enabling the users to tailor the system to their own preferences and daily routines, as well as to modify the system's functionality to suit their own context and evolving feelings towards the system, the perceived utility of the system itself is enhanced, potentially preventing future disengagement by enhancing self-efficacy, user satisfaction and engagement with intelligent systems, resulting in a more user-centric system [148]. The HITL enhances the adaptability by enabling humans to assist intelligent systems in adjusting to evolving/changing conditions, goals or preferences, with advantages in terms of robustness and personalization of the system.
- The **continuous monitoring, evaluation and improvement of the AI actions**, thanks to the integration of human intelligence, promoting a feedback loop for improving the accuracy, system performance and reliability of automated systems. By allowing humans to provide feedback, on the one hand, increases accuracy and enables error detection and correction in cases where the data are noisy, sparse, or ambiguous, and, on the other hand assist the ML model in focusing on relevant or informative data, thus reducing the amount of data that needs processing or labelling. The HITL approach, in other words, is a mean to improve the AI accuracy: the incorporation of the human expertise and judgment into the AI decision-making process supports in identifying and correcting errors, biases, and limitations in AI models. Besides the continuous improvement of AI models through human feedback and fine-tuning,

leading to more accurate and reliable AI systems over time, the human operators is able to offer valuable domain-specific knowledge and contextual understanding that AI algorithms may lack, enabling them to catch and rectify mistakes that might otherwise go unnoticed. The incorporation of human expertise during the training phase can help to guide the learning process, resulting in models that are better aligned with the nuances and complexities of the situation. Similarly, the involvement of human experts in the evaluation phase helps to identify and correct the model errors, leading to more accurate and reliable predictions. The human feedback loop makes possible the refinement of the AI models, ensuring that they capture the relevant patterns and relationships within the data. As regards data, data processing methods based on human-in-the-loop might be useful for data pre-processing, data annotation, and iterative labelling.

Nevertheless, there are also some challenges in integrating the human element into AI-based systems and, more in general, in the HITL approach, ranging from the **potential slowing down of decision-making processes**, due to the time needed for the human intervention, which might reduce the speed and efficiency that automation process, to the **risk of human error** (due to misjudgement, lack of knowledge, or misinterpretation of data) and, in case of human-in- the-loop topic modelling task, some existing HITL techniques might allow **malicious individuals to efficiently train models that serve their purpose**, with possible consequent damages to the society. Another possible lowlight regards the **need for continuous training** for operators interacting with increasingly complex AI systems, to keep them up-to-date and competent in such interaction. Furthermore, the scalability of the AI system might be limited by the need for human intervention, unless the supervisory personnel are proportionately increased, which implies a cost.

It is therefore important to adopt approaches capable of maximizing the benefits of AI while maintaining a significant and manageable role for human intervention. The Explainable AI techniques play a paramount role in this direction, since they make AI-based decision-making processes and outcomes more transparent and understandable, thereby facilitating more informed and efficient HITL, including human oversight.

One of the key aspects characterizing AI-DAPT methodology is that the HITL approach is effectively combined with automation through the underlying use of Explainable AI techniques to explain all AI models, i) supporting the automation of certain pipeline steps or ii) providing the expected prediction/inference results for the problem at hand) and their associated results, or iii) contributing to filtering out (training) data that may hurt the model performance because of poor quality or biases.

The HITL approach is effectively combined with automation through the underlying, implicit application of Explainable AI techniques to explain all AI models (that either support the automation of certain pipeline steps or provide the expected prediction/inference results for the problem at hand) and their associated results. Such techniques also contribute to filtering out (training) data that may hurt the model performance because of poor quality or biases.

As regards **Explainability**, it is a breakthrough in the way to interact with and understand decisions made by the AI systems, making their processes transparent and understandable to humans: it is paramount to adopt a human-centred perspective, providing the human operators with the tools needed for the effective supervision and intervention Likewise, XAI is essential to take into account the human considerations of the human-machine interaction, allowing the users to contribute to improving the AI model. In this direction, it is necessary to make the process more accessible and efficient and integrate explanations into it, giving rise to a bi-directional communication channel between the human operators and the intelligent systems [148], prioritizing the users as the primary driver of this interactive bi-directional process to achieve the desired system behaviour, whilst empowering them to collaborate with intelligent systems towards enhancing the effectiveness of the interaction and the outcomes of the system, ensuring adaptability, which is essential to ensure human comfort, which, in turn, increases trust and acceptance of AI systems. There are several aspects in which XAI adds value to human-machine collaboration. First of all, XAI, providing insights into the "how" and "why" behind decisions made by the AI system, helps the human operators to understand the underlying patterns and decision-making processes, making it easier to identify and correct any errors or biases in the system, as well as facilitating a more effective collaboration between humans and machines, since the operators, thanks to the information provided by XAI, can make informed decisions, taking full advantage of AI systems' data analysis capabilities, and/or provide more accurate feedback for improving and refining the AI models (continuous feedback loop). On the other hand, a better understanding of such patterns and processes improve the trust in the decision of the AI system. However, like for the HITL approach, also the implementation of the XAI might present some challenges. First of all, it might be difficult, in some complex cases, to create adequate explanations, especially for non-expert users. Furthermore, there is a challenge related to the perceived trade-off between AI explainability and performance/accuracy, with the need to seek for a balance, since it is often assumed that in some cases more explainable models can be less accurate than more complex and less transparent ones: in other words, it is often assumed that the pursuit of XAI may come at the cost of sacrificing some degree of accuracy.

Furthermore, there is also the so-called "**illusion of explanation**" [150], occurring when the users find emotional satisfaction in explanations and believe they grasp the processes, ending up with only a superficial understanding of the explanations provided, since such explanations fail to faithfully reflect the realities of decision-making, leading to a false sense of security, misinterpretation, misuse, or over-reliance on explanations. Such illusion of explanation can be originated by different factors, such as the overestimation of human understanding, a phenomenon known as the illusion of explanatory depth (IOED), which gives rise to misinterpretations of explanations, which might be facilitated by the non-transparency and lack of interpretability of the ML systems. However, the users, deployers and developers need to remain vigilant and enabled to discern the actual reasons behind the decisions in question. To mitigate the risk of illusion of explanation, it is key to remind that the explainability is intricately linked to the human-in-the-loop approach to give to the users' comprehensible insights into the outcomes generated by complex models, crafting a comprehensible and reliable explanation, beneficial for both developers and users. Nevertheless, it is important to adopt a cautious approach of the HITL approach itself, rooted in a deep understanding of human-in-the-loop responsibilities and integrating the robustness of probabilistic reasoning with the clarity of logical rules, in order to prevent overreliance and misinterpretation of AI-generated explanations, balancing the need for interpretability and the ability to handle complex, uncertain scenarios. XAI can be directed to distinct groups of people interested in the explanations of the AI system, such as: i) the end-users decision makers, who use the recommendations/predictions of the AI system to make a decision, ii) the affected users, who are people impacted, or potentially impacted, by the recommendations made by an AI system, iii) Regulatory or Supervisory Bodies, which want to check and ensure that the decisions are made in a safe and fair manner, and that society is not negatively impacted by such decisions, iv) AI System Developers, including data scientists, who build or deploy AI systems and need to know if they are working as expected and how to diagnose and improve them.

On the other hand, one of the main challenges posed by some AI systems which the HITL approach and XAI can contribute to prevent or minimize, regards the so-called **"hallucination effect"** [149], consisting in a phenomenon where the AI systems, like text/images generators or neural networks, can generate inaccurate or logically inconsistent responses or data, relying on patterns learned during training that can lead to incorrect or far-fetched conclusions. The hallucination effect occurs when the machine produces an outcome that is not based on real or logical data, but rather on a kind of "digital fantasy"/imaginary elements created by its own learning network. In fact, in these situations the human intervention is paramount to guide, correct and improve decisions made by machines and the XAI is a key element in navigating the complexities and challenges like the hallucination effect. The incorporation of user knowledge into the system, besides being functional to promote the automation of machine learning, can provide training data for machine learning applications and directly accomplish tasks that are hard for computers in the pipeline with the help of machine-based approaches, training an accurate prediction model with minimum cost by integrating human knowledge and experience [147]. The integration of a priori knowledge in the learning framework contributes to deal with sparse data, considering that the learner does not need to induce the knowledge from the data: in case the machine is encouraged to engage with learning human wisdom and knowledge, it would help deal with sparse data, relying on the incorporation of human knowledge and experience into the modelling process to endow machines with more intelligence.

The adoption of the **hybrid science-guided AI model** approach, including both models closer to the science-guided end (the so-called "Machine Learning Complements Science") and models nearer the data-driven end (the so-called "Science Complements ML") [151], based on more automated and less data-intensive methods, frameworks, technologies and solution, is very helpful to prevent the “hallucination effect” and, more in general, to improve the accuracy and explainability of the AI pipelines, to solve complex problems, advance data analysis, prediction and decision-making. By integrating domain knowledge from science into AI frameworks, combining the artificial intelligence techniques with scientific theories and principles, the model's accuracy, interpretability, and predictive power are enhanced, so that the model can make informed predictions and decisions based on both data and established scientific understanding. T3.2 “Hybrid Science-AI Model Explainability & Evaluation within AI Pipelines” is dedicated to design and implement suitable methods and interfaces for serving the explainability and evaluation purposes of an AI pipeline towards end users, in addition to the XAI techniques implicit and embedded in all tasks in WP2-WP3 where AI-based support for the pipeline automation is provided, in order to provide constantly accurate, data-driven and scientifically consistent insights. The hybrid science-guided AI models, despite in general give rise to more accurate and scientifically consistent predictions, might also raise some ethics challenges, such as the introduction of bias into the model and its predictions, due to the incorrect fundamental knowledge in the science-based model: the AI systems might perpetuate or even amplify existing biases present in the data they are trained on, potentially leading to unfair treatment. It’s therefore crucial to ensure that training data are of high quality, diverse and representative. Other ethics challenges of the science-guided AI models, such as those regarding the risk of lack of transparency, the privacy concerns due to the collection and use of large amounts of data to train AI models, which can infringe on individuals' privacy rights, the security issues, being the AI systems exposed to adversarial attacks, where malicious actors manipulate inputs to produce harmful outputs, as well as the risk of misuse and misinformation, since the AI-generated content can be manipulated to spread misinformation or be used maliciously, leading to potential societal harm, are not specific to the hybrid science-guided AI models, but in common with the others AI-based systems.

Algorithmic bias and Risks of discrimination, manipulation, misuse [152]

AI technologies typically rely on algorithms that make predictions to support or even fully automate decision-making. Algorithms, though they can be used for good, they can also negatively affect fundamental rights, such as by violating the right to privacy or leading to discriminatory decision-making. One relevant ethics issue concerns the algorithmic bias, consisting in the systematic and repeatable errors in an AI system that create "unfair" outcomes, in ways different from the intended function of the algorithm. The factors that might generate bias are several, such as the design of the algorithm or the unintended use or decisions relating to the way data is coded, collected, selected or used to train the algorithm. In other words, bias can be introduced to an algorithm in several ways, such as during the creation of a dataset, during the assembling and processing data, as well as bias can emerge as a result of design. In the feedback loop phenomenon, any bias in algorithms can potentially be reinforced over time and exacerbated, leading to results that overestimate realities (the so-called runaway feedback loops), which might bring sensitive ethics concerns in case of “high-risk” AI applications. The feedback loop can be due, for instance, to the data quality. In other cases, the machine learning algorithms tend to put too much weight on training data: a runaway feedback loop might happen, for instance, in machine learning models if they are not controlled for overreacting to small signals and differences in the data. It is therefore paramount for any algorithm development to adopt techniques for avoiding exaggerated predictions, which mirror training data too strongly (so-called overfitting). On the other hand, the quality of the training data and other sources that influence bias and may lead to discrimination need to be regularly assessed by providers and users, in order to avoid that bias and potential discrimination are developed or amplified over time (in case the data based on outputs of algorithmic systems become the basis for updated algorithms). The bias can have impacts ranging from privacy violations to reinforcing social biases of race, gender, sexuality, and ethnicity and the concerns are especially relevant in case of algorithms reflecting "systematic and unfair" discrimination. This kind of bias has been addressed by the GDPR and the AI Act.

The legal frameworks setting requirements to ensure the fairness and trustworthiness of AI were elaborated to overcome the problems associated with AI bias. The EU AI Act includes specific provisions for bias detection, requiring that high-risk AI systems undergo rigorous testing and validation before their deployment in the EU marketplace, including a sets of compliance checks which comprise the mandatory assessment of AI systems’ biases and the handling of any identified risks. Art. 10 “Data and Data Governance” of the AI Act asks for bias examination and mitigation in its section 2 (f-g):

“2. Training, validation and testing data sets shall be subject to data governance and management practices appropriate for the intended purpose of the high-risk AI system. Those practices shall concern in particular: f) examination in view of possible biases that are likely to affect the health and safety of persons, have a negative impact on fundamental rights or lead to discrimination prohibited under Union law, especially where data outputs influence inputs for future operations; (g) appropriate measures to detect, prevent and mitigate possible biases identified according to point (f);”.

As regards the Processing special categories of personal data, Art. 10 “Data and Data Governance” of the AI Act, Section 5 (a-f) states that: “To the extent that it is strictly necessary for the purpose of ensuring bias detection and correction in relation to the high-risk AI systems in accordance with paragraph (2), points (f) and (g) of this Article, the providers of such systems may exceptionally process special categories of personal data, subject to appropriate safeguards for the fundamental rights and freedoms of natural persons. In addition to the provisions set out in Regulations (EU) 2016/679 and (EU) 2018/1725 and Directive (EU) 2016/680, all the following conditions must be met in order for such processing to occur:

- The bias detection and correction cannot be effectively fulfilled by processing other data, including synthetic or anonymised data.
- The special categories of personal data are deleted once the bias has been corrected or the personal data has reached the end of its retention period, whichever comes first.
- The records of processing activities pursuant to Regulations (EU) 2016/679 and (EU) 2018/1725 and Directive (EU) 2016/680 include the reasons why the processing of special categories of personal data was strictly necessary to detect and correct biases, and why that objective could not be achieved by processing other data”.

Furthermore, international standards play an important role in creating AI-based processes that are free from unwanted bias, such as:

- The published ISO/IEC TR 24027:2021 “Bias in AI Systems and AI-Aided Decision Making”, consisting in a technical report with best practices for bias detection and mitigation in AI systems, offering a clear guidance for identifying bias in all stages of an AI system from data collection to deployment. It underlines that most AI systems are prone to the various forms of bias that can distort their fairness and effectiveness because they are based on data and algorithms.
- The forthcoming ISO/IEC 12791 “Bias Mitigation Techniques”, which is expected to provide mitigation techniques that can be applied throughout the AI system life cycle in order to treat unwanted bias.

- IEEE P7003 Standard for Algorithmic Bias Considerations, which is a work-in-progress report, part of the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, aiming to help identify and mitigate unintended, unjustified, and inappropriate biases in algorithmic decision-making systems and addressing various situations where algorithmic bias might occur and provides methodologies for managing these impacts.
- "Towards a Standard for Identifying and Managing Bias in Artificial Intelligence", a document by the National Institute of Standards and Technology (NIST) on how to detect and mitigate bias in AI systems. It identifies three sources of bias: systemic bias, statistical/computational bias, and human bias.

It is recommended that the AI Consortium refers to such standards for avoiding bias to materialize and to address it in case it occurs.

Considering that the AI technologies continue to evolve and the potential for AI systems to perpetuate or exacerbate existing biases, the problem of its detection and mitigation becomes even more important for developing and deploying the AI system in a fair, transparent, and accountable manner and bias detection and mitigation strategies must be effectively integrated into the AI development process from the outset.

8.2.2 Ethical and Legal Reference Framework

The key European legal, regulatory and ethical sources relevant for AI-DAPT system and technological assets, and, thereby, central to the AI-DAPT Trustworthy Framework can be classified into two main areas, respectively pertaining to Artificial Intelligence and on Data. They are listed in the following table, which is an update of the one inserted in D6.1 "Dissemination, Communication, Engagement and Innovation Plan", at the beginning of the project.

Artificial Intelligence
AI Act (AIA), Regulation (EU) 2024/1689 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 13 June 2024, laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). It is the first-ever legal framework for AI and, building on the Commission White paper on AI, it moves forward towards trustworthy and ethical AI systems in the European market, with a balanced approach to innovation, safety, security, and privacy. A provisional agreement was reached on December 9th 2023 and includes safeguards for general-purpose AI, limitations or bans for AI applications, like social scoring, manipulation, biometrics, etc. After the approval process, it will become applicable law reasonably during the summer of 2024. Some provisions of the AI Act are particularly relevant for the AI-DAPT Trustworthy Framework [153]: • **Risk-based approach and classification of the AI systems according to their risks**. The AI DAPT considers the AI applications according to a typology of risks from: 1. **Unacceptable risk AI systems**: they imply harmful uses of AI, contravening the EU values, and are therefore banned, with some exceptions. 2. **High risk AI systems**: they have a negative impact on fundamental rights and safety. For this kind of applications, several mandatory requirements (including a conformity assessment) are provided. All high-risk AI systems will be assessed before going to the market and throughout their lifecycle. The High-risk AI systems (Chapter III) are those (Art. 6):

i. Used as a safety component or a product covered by EU laws in Annex I and required to undergo a third-party conformity assessment under those Annex I laws; or

ii. those under Annex III use cases (below), except if: i) the AI system performs a narrow procedural task; ii) improves the result of a previously completed human activity; ii) detects decision-making patterns or deviations from prior decision-making patterns and is not meant to replace or influence the previously completed human assessment without proper human review; or iv) performs a preparatory task to an assessment relevant for the purpose of the use cases listed in Annex III;

iii. AI systems are always considered high-risk if they profile individuals, i.e. automated processing of personal data to assess various aspects of a person's life, such as work performance, preferences, reliability, behaviour, location or movement.

In case the providers whose AI system falls under the use cases in Annex III believe it is not high-risk, they have to document such an assessment before placing it on the market or putting it into service.

3. **Limited risk AI systems**: for instance, AI systems that generate or manipulate image, audio or video content. For them, a limited set of obligations (e.g. transparency) are provided. The developers and deployers must ensure that end-users are aware that they are interacting with AI.

4. **Minimal risk AI systems**: these are all the other AI systems and are the majority of AI applications currently available on the EU single market. They can be developed and used in the EU without additional legal obligations (besides those posed by the existing legislation).

- **The majority of obligations fall on providers (developers) of high-risk AI systems**. The AI Act (Art. 8–17) sets forth the requirements for providers of high-risk:

1. Establish a risk management system throughout the high-risk AI system's lifecycle.

2. Conduct data governance, ensuring that training, validation and testing datasets are relevant, sufficiently representative and, to the best extent possible, free of errors and complete according to the intended purpose.

3. Draw up technical documentation to demonstrate compliance and provide authorities with the information to assess that compliance.

4. Design their high-risk AI system for record-keeping to enable it to automatically record events relevant for identifying national level risks and substantial modifications throughout the system's lifecycle.

5. Provide instructions for use to downstream deployers to enable the latter's compliance.

6. Design their high-risk AI system to allow deployers to implement human oversight.

7. Design their high-risk AI system to achieve appropriate levels of accuracy, robustness, and cybersecurity.

8. Establish a quality management system to ensure compliance.

- **The deployers of the AI system are considered as users**: in fact, the users are natural or legal persons deploying an AI system in a professional capacity. The users are not the affected end-users. In particular, the users (deployers) of high-risk AI systems have some obligations, though less than providers (developers).

- **Specific provisions and requirements apply to the General-purpose AI (GPAI)**: The GPAI model providers must provide technical documentation, instructions for use, comply with the Copyright Directive, and publish a summary about the content used for training. Free and open licence GPAI model providers only need to comply with copyright and publish the training data summary, unless they present a systemic risk. In case of GPAI models that present a systemic risk – open or closed, the providers must also conduct model evaluations, adversarial testing, track and report serious incidents and ensure cybersecurity protections.

  On 12 July 2024, the AI Act was published in the EU's Official Journal and took effect on 1 August 2024. From the date of the entry into force, the following milestones will follow according to Article 113:

  1. 6 months later – Chapter I and Chapter II (prohibitions on unacceptable risk AI) will apply.
  2. 12 months later – Chapter III Section 4 (notifying authorities), Chapter V (general purpose AI models), Chapter VII (governance), Chapter XII (confidentiality and penalties) and Article 78 (confidentiality) will apply, with the exception of Article 101 (fines for GPAI providers).
  3. 24 months later – The remainder of the AI Act will apply, except.
  4. 36 months later – Article 6(1) and the corresponding obligations in this Regulation will apply.
  5. Codes of practice must be ready 9 months after entry into force according to Article 56.

  The AI Act will be updated, amended and implemented through implementing acts and delegated acts. The former tends to focus on implementation of the act (such as by providing official guidance on compliance). The latter are closer to legislative amendments, changing details written into the AI Act. Both are powers given to the Commission to update the act in response to technological developments, as well to provide non-essential details at a later date. Furthermore, the AI Office is expected to continue bringing expertise to the EU and advise on some implementing and delegated acts, as well as on many other areas where expertise might be needed during implementation and enforcement. Pending the formal adoption, the AI Pact was adopted, which anticipated the implementations of some AI Act requirements with voluntary companies.

AI Liability Directive (AILD) Proposal, COM (2022) 496 final "Proposal for a Directive of the European Parliament and of the Council on adapting non- contractual civil liability rules to artificial intelligence". This instrument lays down uniform rules for certain aspects of non-contractual civil liability for damage caused with the involvement of AI systems for ensuring that persons harmed by AI systems enjoy the same level of protection as persons harmed by other technologies. Contrary to some rumours regarding the assertion that the AI Liability Directive would be abandoned, the EC reaffirmed its commitment to advancing the Directive, as shown by the recent dissemination of the amended proposal to EU governments and lawmakers for further examination. It sets for the first time a targeted harmonisation of national liability rules for AI for making it easier for victims of AI-related damage to get compensation and to ensure that victims benefit from the same standard of protection across the EU when harmed by AI products or services, alleviating the burden of proof for damages caused by AI systems and ensuring an easier access to redress for the victims and provides broader protection for victims (individuals or businesses), whilst increasing guarantees. It states: i) the right of access to evidence: subject to certain conditions, a court (or, in limited circumstances, third parties) can order to a provider of a high-risk AI system to disclose relevant and necessary evidence about their product. In addition, the victims have the right of access to evidence from companies and suppliers when high-risk AI is involved; ii) a rebuttable presumption of causality, when a relevant fault has been established and a causal link to the AI performance seems reasonably likely. Subject to certain conditions, national courts will presume, for the purposes of applying liability rules to a claim for damages, that the output produced by the AI system (or the failure of the AI system to produce an output) was caused by, for example, the fault of the AI provider. In this way, the providers of AI systems are responsible and, in some cases, the user of AI systems (each as defined in the EU AI Act). However, there is the right to fight a liability claim based on a presumption of causality.

Revised Product Liability Directive (RPLD) Proposal, COM (2022) 495 final, “Proposal for a Directive of the European Parliament and of the Council on liability for defective product”. Also, this instrument is aimed to properly address the needs of the digital age, circular economy business models and global value chains, renovating the existing Product Liability Directive (adopted in 1985). It addresses liability for products such as software (including artificial intelligence systems) and digital services, affecting how the product works (e.g. navigation services in autonomous vehicles), providing the companies with legal certainty and ensuring that victims get fair compensation when defective products cause harm. It alleviates the burden of proof for victims under certain circumstances and recognize the liability rules for companies that substantially modify products before resale to extend the product lifecycle (circular economy). The recoverable damages comprise not only personal injury, death and damage to personal property, but also loss or corruption of data and medically recognized harm to psychological health. The non-exhaustive list of factors to consider in assessing defect includes also, for instance, self-learning abilities.
Among its provisions, the following are particularly relevant:

- The compensation for damage is allowed also when products like robots, drones or smart-home systems are made unsafe by software updates, AI or digital services that are needed to operate the product, as well as when manufacturers fail to address cybersecurity vulnerabilities.

- Alleviation of the burden of proof for victims in complex cases, such as those involving AI (as highlighted in the accompanying document), where it would be difficult for the victim to prove liability pursuant to the RPLD, there are five scenarios in which the causal link between defectiveness and damage is impossible to prove due to the technical or scientific complexity of the product: in such cases it is presumed. One of such scenarios is aimed to prevent the so-called ‘black box’ effect of AI systems: in the given circumstances, the claimant will only need to prove that the AI at hand contributed to the damage and that the product is likely to be defective.

- The time limit for bringing claims is still 3 years (from the earlier of either the day on which the claimant becomes aware, or should reasonably have become aware, of the damage, the defect and the identity of the relevant economic operator who is liable). However, the liability period expires after ten years since the defective product was placed into the market or, in case the injury is not immediately apparent, after 15 years.

- The existing strict liability (i.e. no fault) regime for defective products across the EU extended as regards the scope of claims that can be brought and the range of damages that can be recovered: for instance, it is also applicable to the defects resulting from cybersecurity risks, connectivity risks, software updates (or lack of updates), with limited exceptions for software updates beyond a manufacturer's control, e.g. due to the user not installing the update. On the other hand, it is simplified for consumers to prove their case.

- The recoverable damages now include loss or corruption of data and medically recognised harm and psychological health.
- The non-exhaustive list of factors to take into account in assessing defect is enlarged and includes, for instance, product safety requirements (including safety-relevant cybersecurity requirements), foreseeable misuse and self-learning abilities.

The **AI innovation package** to support Artificial Intelligence startups and SMEs [154], adopted in January 2024. It consists of a set of measures to support European startups and SMEs to develop trustworthy AI, respectful of EU values and rules. It also includes the amendment of the EuroHPC Regulation to set up AI Factories, expected to be paramount within the EU's supercomputers Joint Undertaking activities with provisions, for instance, on AI-dedicated supercomputers to enable fast machine learning and training of General Purpose AI (GPAI) models. The European AI Start-Up and Innovation Communication foresees additional key activities, such as the "GenAI4EU" initiative, aiming to support the development of novel use cases and emerging applications in Europe's 14 industrial ecosystems (in diversified application areas, such as robotics, biotech, health, manufacturing and mobility), as well as the public sector.

Ethics Guidelines for Trustworthy AI and ALTAI Assessment List. Both of them were prepared by the High-Level Expert Group on Artificial Intelligence (AI HLEG), appointed by the EC in 2018. Their objective is to foster an ethical, trustworthy approach to AI, functional to enable responsible and sustainable AI innovation in Europe. The Ethics Guidelines were published in 2019. They are not legally binding such Guidelines do not offer advice on legal compliance for AI. They describe ethical principles relevant to build a trustworthy AI, which must display the following three characteristics:

- **Lawfulness**, relying upon the "human-centric approach" to AI, where fundamental human rights are deemed as the foundation of Trustworthy AI. In this direction, the EU Charter and European Convention of Human Rights are considered the key for any legislative source in the field of AI.
- **Robustness**, considering the ability of AI to operate in any situation, especially if unpredictable events or malicious attacks occur.
- **Ethically-soundness**, requiring that technological design, development and use of AI are compliant with the EU ethical values listed in the Guidelines themselves.

The Guidelines identify ethical principles governing AI on the basis of fundamental human rights and translate them into seven requirements that AI systems must fulfil in order to be considered trustworthy:

1. **Human agency and oversight**: AI system must be supportive to human action, human autonomy and decision-making. They have to act "as enablers to a democratic, flourishing and equitable society by supporting the user's agency and foster fundamental rights, and allow for human oversight" [155]. AI systems must promote fundamental rights, benefitting people, reducing risk of infringement on such rights in order to respect the rights and freedoms of others. Any kind of unfair manipulation, deception, herding, diminishing, limiting, or misleading human autonomy and/or conditioning must be avoided. The principle of user autonomy must be central to the AI system's functionality. It is paramount ensuring that the AI technology does not undermine human autonomy or causes other adverse

effects: thereby human oversight must be allowed, through governance mechanisms (such as a human-in-the-loop, human-on-the-loop or human-in-command approaches) and in varying degrees, taking into account the application area and the potential risk of the AI solution.

2. **Technical robustness and safety**: safe, reliable algorithms must be in place and must be capable to handle errors or inconsistencies during all phases of the AI systems' life cycle. This is closely linked to the principle of prevention of harm and requires a preventative approach to risks in AI systems' development, so that to reliably behave as intended, whilst minimizing unintentional and unexpected harm. It is important to consider the potential changes in the AI system's operating environment or the presence of other agents (human and artificial) potentially interacting with the system in an adversarial manner. The AI system must ensure the physical and mental integrity of humans and must be resilient to attack and security, with safeguards for a fallback plan in case of problems.

   Depending on the magnitude of the risk posed by an AI system and on the application area, appropriate level of safety measures must be ensured. The AI system must also be able to make correct judgements, for example to correctly classify information into the proper categories, ensuring accuracy. Its results must be reproducible, exhibiting the same behaviour when repeated under the same conditions, as well as reliable, working properly with a range of inputs and in a range of situations and preventing unintended harms.

3. **Privacy and data governance**: in compliance with the GPDR and in line with the principle of prevention of harm, citizens should have full control of their data. The privacy is a fundamental right particularly affected by AI systems. The data must not be used against the citizens or in any discriminatory way. Adequate data governance mechanisms must be adopted, covering the quality and integrity of the data used, their relevance in the specific case, their access protocols and data processing in a manner that guarantee privacy and data protection throughout the AI system's entire lifecycle. The personal data collected must not be used to unlawfully or unfairly discriminate against the data subjects. The quality and integrity of the datasets must be guaranteed, without biases, inaccuracies, errors and mistakes, especially in case of self-learning systems. Proper data protocols governing data access must be foreseen and respected, outlining who can access data and under which circumstances.

4. **Transparency**: It is key to ensure the traceability and explainability of the AI systems, encompassing also transparency of elements relevant to an AI system (the data, the system and the business models). The traceability implies that the datasets, processes and decision of the AI system's decision must be documented to the best possible extent. The explainability refers to the ability of the AI system to explain both the its technical processes and the related human decisions (e.g. application areas of a system): the decisions made by an AI system can be understood and traced by human beings. It must be clear to the humans that they are interacting with an AI system and it must be identifiable as such.

5. **Diversity, non-discrimination, and fairness**: the AI system's lifecycle must ensure inclusion, diversity, equal treatment and access. The inclusive design processes must be followed, avoiding unfair bias, including inadvertent historic bias, incompleteness and bad governance models which could give rise to unintended (in)direct prejudice and discrimination or exacerbate prejudice and marginalization. This is related with the principle of fairness and also pertains to the way in which AI technology is developed: it is key to avoid unfair bias by putting in place proper oversight mechanisms in relation to the system's purpose,

constraints, requirements and decisions, as well as to adopt a user-centric approach and to follow the universal design principles. Efforts must be directed to allow to all people to use the AI service, regardless of their age, gender, abilities or characteristics, avoiding one-size-fits-all approach. Stakeholder involvement should be encouraged for this purpose.

6. **Societal and environmental wellbeing**: the sustainability and ecological responsibility of AI systems must be promoted at the maximum extent, as well as measures strengthening the environmental friendliness of AI systems' supply chain. Likewise, the positive social impact and enhancement of social skills must be fostered with a wide perspective taking into account possible effect on democracy and society at large.

7. **Accountability**: the responsibility and accountability for AI systems and their outcomes, both before and after their development, deployment and use, must be ensured. Mechanisms must be put in place on this purpose, including auditability. Auditability entails the enablement of the assessment of algorithms, data and design processes (paying attention to safeguard the business models and intellectual property related to the AI system). It also refers to the evaluation by internal and external auditors with the preparation of evaluation reports, as well as the ability to identify, assess, document and minimize the potential negative impacts of AI systems. The use of impact assessments, such the Algorithmic Impact Assessment, both prior to and during the development, deployment and use of AI tool, is recommended, in a proportionate manner in relation to the risk that the AI systems pose. In case of trade-offs, a rational and methodological approach must be used to tackle with them, explicitly acknowledged and assessed any risk to ethical principles and fundamental rights, with the limit of ethically acceptability. In case of unjust adverse impact, adequate redress mechanisms must exist and be applied.

The "Assessment List for Trustworthy Artificial Intelligence" (ALTAI) for self-assessment, supports their actionability, by translating them into an accessible and a dynamic checklist.

In AI-DAPT, special attention will be given on technical robustness and safety, human agency and oversight, privacy/IPR preservation and data governance, transparency, traceability, diversity, non-discrimination, and fairness.

Living guidelines on the responsible use of Generative AI in research [156]**.** Guidelines developed by the European Research Area on the use of generative AI in research for funding bodies, research organisations and researchers, both in the public and private research ecosystems.

These guidelines, which are not binding, focus on the generative Artificial Intelligence, setting out common directions on its responsible use, and aim prevent misuse and ensure that Generative AI plays a positive role in improving research practices, by setting out common directions on the responsible use of generative AI. On the basis of key principles on research integrity and on already existing frameworks for the use of AI in general and in research specifically, and building on the commonalities of the currently emerging guidelines from various stakeholders, they identify key principles for the responsible use of generative AI in research, namely reliability in ensuring the quality of research, reflected in the design, methodology, analysis and use of resources, the honesty in developing, carrying out, reviewing, reporting and communicating on research transparently, fairly, thoroughly and impartially; Respect for research participants, research subjects, colleagues, society, ecosystems, cultural heritage and the environment; Accountability for the research in its entire lifecycle, including responsibility for all the output of a research, underpinned by the notion of human agency and oversight.

Data

Data Governance Act (Regulation (EU) 2022/868), which is already applicable, is functional to oversee the reuse of publicly or protected data across various sectors, facilitating data sharing by the data intermediaries and promoting data sharing for altruistic reasons and enhancing trust in the sharing and reuse of data.

Data Act (Regulation EU 2023/2854), which entered into force on 11 January 2024, is directed to harness the potential of the ever-increasing amount of industrial data, in order to benefit the European economy and society. It complements the Data Governance Act by ensuring fairness in the allocation of the value of data amongst stakeholders, in view of creating a European single market for data in which data can flow between sectors and Member States in a safe and trusted manner for the benefit of the economy and society. It entered into force on 11 January 2024 and will become applicable on 12 September 2025. The European Commission plans to recommend by autumn 2025 a set of model contractual terms to help businesses conclude data-sharing contracts that are fair, reasonable and non-discriminatory (Chapters II and III of the Data Act) and will also provide guidance on reasonable compensation and the protection of trade secrets.

Considering that the rapid growth in the availability of products connected to the internet (‘connected products’, which together compose a network known as the Internet-of-things, IoT), significantly increase the volume of data available for reuse in the EU, it is paramount to ensure their availability and use. In this direction, the Data Act aims to make data (in particular industrial data) more accessible and usable, i) encouraging data-driven innovation and increasing data availability, ii) ensuring fairness in the allocation of the value of data amongst the different actors and iii) clarifying who can use what data and under which conditions.

Under the Data Act, users of connected products (businesses or individuals that own, lease or rent such a product) have a greater control over the data they generate. At the same time, incentives are maintained for those who invest in data technologies and there are situations where a business has a legal obligation to share data with another business.

The Data Act is without prejudice to the laws on the protection of intellectual property rights as well as it is fully compliant with the GDPR.

GDPR, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). The GDPR is a comprehensive regulatory framework laying down provisions for ensuring that personal data enjoys a high standard of protection everywhere in the EU and for giving individuals back control over of their personal data.

These **GDPR definitions and concepts** are particularly relevant to AI-DAPT:

- **Data subject**: “identified or identifiable natural person[s]”. Only natural persons (human beings) are beneficiaries of the data protection rules.

- **Personal data**: data relating to an identified or identifiable person (the “data subject”). They concern information about an individual whose identity is either manifestly clear or can be established from additional information. All reasonable means that are likely to be used to directly or indirectly identify the natural person need to be considered, being the GDPR applicable if the person concerned is identifiable, in a direct or indirect way. The special categories of personal data outlined by the Art 9 (the so-called “sensitive data”) need enhanced protection: personal data revealing racial or ethnic origin; personal data revealing political opinions, religious or other beliefs, including philosophical beliefs; personal data revealing trade union membership; genetic data and biometric data processed for the purpose of identifying a person; personal data concerning health, sexual life or sexual orientation.

- **Data processing**: "'processing of personal data' [...] shall mean any operation [...] such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction" of personal data. The automated data processing refers to the operations performed on "personal data wholly or partly by automated means";
- **Users of the personal data**, include: i) the "Data Controller", determining the means and purposes of processing the personal data of others. If several persons take this decision together, they may be 'joint controllers'; the "Data Processor" (natural or legal person), processing personal data on behalf of a controller; the "Recipients", the person to whom personal data are disclosed; "third party": a natural or legal person (other than the data subject, the controller, the processor and persons who are authorised to process personal data under the direct authority of the controller or processor.

- **Anonymisation**, consisting in the process allowing that all identifying elements are eliminated from a set of personal data so that the data subject is no longer identifiable. No element may be left in the information which could, by exercising reasonable effort, serve to re-identify the person(s) concerned.

- **Pseudonymisation**, consisting in the process of removal from the personal information any attributes (name, date of birth, sex, address, or other elements) that could lead to identification and their replacement by a pseudonym. For the GDPR it is "the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person". The pseudonymised data are still personal data and are therefore subject to the GDPR and other data protection rules.

The **key GDPR principles set forth by art. 5 and governing the processing of personal data** will be followed by the Consortium in the design, development and deployment of AI-DAPT technology and its application within its experiments. They are the following:

- **Lawfulness, fairness and transparency**, which requires that the lawfulness of the processing requires either the consent of the data subject or other lawful basis (necessity to enter a contract; a legal obligation; necessity to protect the vital interests of the data subject or of another person; necessity for performing a task in the public interest; necessity for the legitimate interests of the controller or a third party, if they are not overridden by the interests and rights of the data subject). The AI-DAPT Consortium will rely on the processing of personal data for research purposes, besides on informed consent, regarding the contact details and opinions/personal data of volunteers participating to the testing activities in the pilots.

- **Purpose limitation**, according to which, any processing of personal data must be done for a specific well-defined purpose and only for additional, specified, purposes that are compatible with the original one.

- **Data minimization**, which states that the processing of personal data will be limited to what is strictly necessary to fulfil the purpose of the processing.

- **Data accuracy**, asking that all processing operations inaccurate data are erased or rectified without delay and that data are checked regularly and kept up to date to secure accuracy.

- **Storage limitation**, which requires that the will delete or anonymize the personal data as soon as they are no longer needed for the purposes for which they were collected. However, the lawful storage of data which are no longer needed could happen throughout their anonymization. Personal data may be retained for up to five years in order to comply with auditing constraints.

- **Security, integrity and confidentiality**, asking to implement the appropriate technical or organisational measures when processing personal data to protect the data against accidental, unauthorised or unlawful access, use, modification, disclosure, loss, destruction or damage". The appropriateness has to be determined by taking into account "the state of the art, the costs of implementation and the nature, scope, context and purpose of processing, as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons".

- **Accountability**, according to which, the compliance of the processing operations with the GDPR requirements has to be ensured and appropriate measures have to be taken to promote and safeguard data protection in the processing activities.

Furthermore, in AI-DAPT it is important to respect the **data subjects' rights** and ensure their exercise, as set forth by art. 13 and following of the GDPR: right to information, right of access, right of rectification, Right to Erasure or Right to be forgotten, Right to Restriction of Processing, Right to Data Portability, Right to Object and Right in relation to automated decision-making and profiling.

Regulation on the free flow of non-personal data (Regulation 2018/1807 on a framework for the free flow of non-personal data in the European Union). This regulation lays down rules applicable to any kind of data other than personal data and is directed to give rise to a harmonized approach to the free movement and portability of data in the EU, as well as to improve legal certainty and create a level playing field for all market players. It complements the GDPR provisions in aspects related to non-personal data within the Digital Single Market, creating legal certainty for businesses to process their data wherever they want in the EU, whilst raising trust in data processing services and countering vendor lock-in practices.

This instrument acknowledges the relevance of data for business processes in companies of all sizes and in all sectors, as well as the opportunities which the new digital technologies are opening up.

In case of mixed datasets (including personal and non-personal data), the Free Flow of Non-Personal Data Regulation applies to the non-personal data part of the dataset, whilst the GDPR applies to the personal data part of the dataset. In case of inextricably linked non-personal data part and the personal data parts, the rights and obligations provided by the GDPR fully apply to the whole mixed dataset. In addition, A Practical guidance [157] for businesses on how to process mixed datasets was provided by the EC, including practical examples dwelling upon the rules to follow in these situations. Among its provisions, it is interesting the one setting incentives for industry to develop self-regulatory codes of conduct on the switching of service providers and the porting of data (supported by the European Commission).

No obligations are set on businesses to decide where their data are to be processed. Furthermore, both the GDPR and art. 4 (1) of this Regulation prohibits the data localization requirements[1], restricting the free movement of non-personal data within the EU, and without prejudice to already existing restrictions laid down by EU law and unless they are justified on grounds of public security in compliance with the principle of proportionality. Therefore, the measures restricting the free movement of data within the EU may be set out in laws, in administrative regulations and provisions or even result from general and consistent administrative practices.

The Regulation also promotes the data portability between businesses, avoiding vendor lock-in practices, occurring when users cannot switch between service providers because their data are "locked" in the provider's system and cannot be transferred outside of the vendor's IT system. This provision is directed to make it easier to port data from one IT environment to another one (i.e. either to another provider's systems or to on-site systems), thereby fostering competition between service providers.

e-Privacy Directive (Directive 2002/58/EC on privacy and electronic communications, replacing the Directive 97/66/EC and partially amended by Directive 2009/136/EC) and **e-Privacy Regulation Proposal** COM (2017) 10 final 2017/0003, "Proposal for a Regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC.

The former states provision, complementing the GDPR in this domain, for the processing of personal data and the protection of privacy in the sector of electronic communications, telecommunications networks and internet services. In relation to AI-DAPT, the most relevant articles regard the security of networks and services, the confidentiality of communications, the access to stored data, the processing of traffic and location data.

The EU member states transposed the ePrivacy Directive into their national legal frameworks. It is applicable to entities providing electronic communication services in the EU and therefore it might be relevant to AI-DAPT in a direct or indirect way. In particular, in AI-DAPT the provisions on the following aspects should be taken into account: traffic data and location data (art. 2), the obligation of adopting security measures appropriated to the risk (art 4), the protection to confidentiality of the communications among individuals (art 5), the user's consent (art.6) and data retention (art. 15).

The focus is on the confidentiality of electronic communications, consent requirements and on the protection of online privacy in the electronic communications sector. One of its main aspects regards the cookies, whilst it is necessary to gain the user's consent after the provision with information about the purpose of the data storage and the opportunity to accept or opt-out. The providers of electronic communication services are asked to ensure that their services are secure and inform their users whenever a risk (for instance of a data breach) might leave their personal data vulnerable to misuse. In relation to personal data, it states that the providers of services must erase or anonymize personal data, when there is no longer need of them. Except in specific situation, the personal data may only be retained upon user's consent. The location data obtained through electronic communications must be processed with informed consent and should be anonymized when no longer needed. Appropriate security measures to safeguard users' data must be put in place by the providers of electronic communication services and they have to inform the users and the relevant authorities in case of a security breach involving personal data. Rules are also set forth

[1] According to the Article 3(5) of the Regulation, the data localization requirements are "any obligation, prohibition, condition, limit or other requirement provided for in the laws, regulations or administrative provisions of a Member State or resulting from general and consistent administrative practices in a Member State and in bodies governed by public law, including in the field of public procurement, without prejudice to Directive 2014/24/EU, 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

on how traffic data, which includes information about communication between individuals, can be processed and stored. The ePrivacy Directive, like the GDPR, requires consent to collect and process personal data, thought the GDPR also outlines other principles of lawful processing Both the instruments require robust security measures to protect user's information. The EC adopted the proposal of Regulation on Privacy and Electronic Communications (E-privacy Regulation), with the aim of modernizing this framework, leaving untouched the objectives and principles. The proposed new legal framework, besides being better aligned with the GDPR, addresses new challenges to privacy. However, it is not clear when it will enter into force.

Open Data Directive (including High Value Datasets). Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast). It applies to the open data and the re-use of public sector information, laying down common rules for a European market for government-held data for making public sector and publicly funded data re-usable, building around two key strands of the internal market: transparency and fair competition. It replaced the Public Sector Information (PSI) Directive.

Commission Implementing Regulation (EU) 2023/138 of 21 December 2022 laying down a list of specific high-value datasets and the arrangements for their publication and re-use. In January 2023, the Commission published a list of specific high-value datasets by way of an implementing act. The public sector bodies have made them available for re-use, free of charge, within 16 months.

The re-use of high-value datasets is associated with important benefits for the society and economy. These datasets are subject to a separate set of rules ensuring their availability free of charge, in machine readable formats accessible via an Application Programming Interface (API) and, where relevant, as bulk downloads.

The thematic categories of high-value datasets listed by Article 13(1) of the Directive also includes domain potentially relevant to AI-DAPT, namely "companies and company ownership", "earth observation and environment", "statistics" and "mobility". The rules are intended to make these datasets available to fuel artificial intelligence and data-driven innovation.

NIS 2 Directive EU 2022/2555). It is directed to achieve a high common level of cybersecurity across the European Union, modernising the existing legal framework to keep up with increased digitisation and an evolving cybersecurity threat landscape. This instrument became enforceable as of 16 January 2023. By 17 October 2024, Member States must adopt and publish the measures necessary to comply with the NIS 2 Directive.

Its legal measures to boost the overall level of cybersecurity in the EU will ensure, among others, a culture of security across sectors that are vital for our economy and society and that rely heavily on ICTs, including sector relevant for AI-DAPT, such as energy, healthcare and digital infrastructures.

The operators of essential services in such sectors will have to take appropriate security measures and notify relevant national authorities of serious incidents. Key digital service providers will have to comply with the security and notification requirements under the Directive.

Cybersecurity Act 2019/881/EU, as well as its proposed amendment of 18 April 2023.The EU Cybersecurity Act has two main objectives: i) strengthening the mandate of the European Union Agency for Cybersecurity (ENISA), which is responsible for ensuring safe internet practices within the EU and ii) establishing an EU-wide cybersecurity certification framework for ICT products, services and processes, setting cybersecurity standards and guidelines. The proposed amendment (April 2023) is directed to enable, through Commission implementing acts, the adoption of European cybersecurity certification schemes for "managed security services", in addition to ICT products, services and processes, which are already covered under the Cybersecurity Act. This framework gives an opportunity for businesses supplying digital products, services and processes (or providing managed security services) to market them certified as meeting EU cybersecurity standards. The certification will be on a voluntary basis. This Act proactively addresses growing risks and threats related to data security, besides offering a unified approach to protect digital information and systems that handle and transfer these data and ensuring high levels of security in products and services. These measures are expected to instil greater consumer trust and efficaciously protect the overall cyber environment in Europe, fostering digital resilience and a collective response to potential cyber threats.

European Data Strategy (COM 2020 66 final, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions “A European strategy for data”. It is one of the cornerstones of the EU’s digital strategy for creating a solid data-driven economy. It is the enabling legislation for the development of common European data spaces and is directed to promote the creation of a single market for data relying on data sovereignty, ensuring a wider availability of data for use in the economy and society, whilst paying attention to keep the companies and individuals who generate the data in control.

IDSA Rulebook 2023, concerning the IDS Data Sovereignty paradigm. This paradigm id directed to help in building trust in data sharing thanks to the technological enforcement of contractual provisions for enabling the data providers to keep a certain control and self-determination over the reuse of the data they provide.

Human Rights

European Convention on Human Rights, adopted in 1950 and the **Charter of Fundamental Rights of the European Union**, 2016/C 202/02.

These sources are based on the milestone document in the history of human rights (Universal Declaration of Human Rights, 1948). They enshrine into EU law a wide set of fundamental rights enjoyed by EU citizens and residents. They are complemented by the European Court of Human Rights’ jurisprudence, which is a useful tool for their interpretation.

In addition to this, it is important to consider the “**European Declaration on Digital Rights and Principles**”, signed by the European Commission Ursula von der Leyen on 15 December 2022.

Such Declaration shows the EU's commitment to a secure, safe and sustainable digital transformation that puts people at the centre, in line with core EU values and fundamental rights: in other words, it acknowledges that the European values, as well as the rights and freedoms enshrined in the EU's legal framework, must be respected online as they are offline.

It provides a reference framework for citizens on their digital rights and for the companies a guidance on how to deal with new technologies, so to get the most out of the digital transformation in Europe.

The array of digital rights and principles outlined in it are meant to complement existing rights, rooted in the Charter of Fundamental Rights of the EU, as well as the data protection and privacy legislation.

The Rights and principles that have to guide the digital transformation, regard the following aspects:

- **People at the centre**: the digital technologies should protect people’s rights, support democracy, and ensure that all digital players act responsibly and safely.

- **Freedom of choice**: the individuals should benefit from a fair online environment, be safe from illegal and harmful content. Besides, they should be empowered when they interact with new and evolving technologies like artificial intelligence.

- **Safety and security of the digital environment**, ensuring the protection and empowerment of all the users, irrespective for instance of their age and socio-cultural background.

- **Solidarity and inclusion**, so that technology can unite, not divide, people.
- **Participation**, so that citizens are able to engage in the democratic process and have control over their own data.
- **Sustainability and green transition**, supporting them via digital devices, with people's awareness on the environmental impact and energy consumption of their devices.

These sources are relevant in AI-DAPT, considering its human-centric approach and trustworthy framework. In particular, the following features of the Declaration are aligned with AI-DAPT ethics approach and overall Trustworthy Framework: i) putting people at the centre of the digital transformation; ii) increasing safety, security and empowerment in the digital environment; iii) promoting sustainability.

The legal and ethical reference framework relevant for AI-DAPT might be further enriched at a later stage, also taking into account the future regulatory developments. In section 8.3 it is integrated with demonstrator-specific relevant regulatory and ethical sources, which were considered for eliciting the legal and ethical requirements specific to them.

8.2.3 Ethical and Legal Requirements for AI-DAPT Framework

AI-DAPT Trustworthy Framework is guided by the Ethics-by-Design Approach, which is functional to help the technical team to reflect on and address any potential ethics concerns during the design and development phase, so to prevent at the maximum extent ethical issues from arising at a later stage.

The Ethics-by-Design approach adopted by the Consortium is described in Sect. 1 of this document and relies on the guidance document "Ethics of Use Approaches for Artificial Intelligence" [158].

Following such approach, the legal mandates and the ethical principles should be taken into account by the development team and the subsequent ethical and legal requirements should be met, considering them as system requirements, together with the other technical, functional and non-functional requirements.

On the bases of the ethics principles, outlined in the guidance document mentioned above and applicable to the AI-DAPT Framework, enriched with those stemming from the Ethics Guidelines for Trustworthy AI and the ALTAI Assessment List, as well as on the basis of the applicable legal framework, with special attention to the AI Act and other sources in the AI domain and in the data domain (including those underway, such as the AILD), the legal and ethical requirements for AI-DAPT system have been elicited.

They are described in the following table:

AI-DAPT Ethical and Legal Requirements
R1. Respect for Human Agency and Autonomy
• All the people affected by the AI-DAPT system, including the end-users and the system operators must maintain the ability to make basic decisions and keep basic freedoms, without being subordinated, coerced, deceived, manipulated, objectified, or dehumanized. • Comprehensible information about the logic involved by the system must be provided to them, as well as about the significance and the envisaged consequences for them. It should be avoided that the AI-DAPT AI-system generates confusion for some or all end-users or

subjects on whether a decision, content, advice or outcome is the result of an algorithmic decision. The end-users or other subjects must be adequately made aware that a decision, content, advice or outcome is the result of an algorithmic decision.

- The AI-DAPT system and/or tools must not take decisions autonomously and without human oversight and possibilities for redress, when fundamental personal issues or economic, social and political issues are concerned.

- The disproportionate attachment, over-reliance or the addiction to the system and its operations should be prevented. Adequate procedures must be adopted to avoid that end-users over-rely on the system.

- The system operators and, as much as possible, end-users should maintain the ability to control, direct and intervene in basic operations of the system and/or tools.

R2. Human Oversight

- Appropriate oversight measures must be in place, through governance mechanisms such as:
  1. Human-in-the-loop (HITL) approach, which refers to the capability for human intervention in every decision cycle of AI-DAPT.
  2. Human-on-the-loop (HOTL) approach, which refers to the capability for human intervention during the design cycle of the system and monitoring the system's operation.
  3. Human-in-command (HIC) approach, which refers to the capability to oversee the overall activity of the AI system (including its broader economic, societal, legal and ethical impact) and the ability to decide when and how to use the same in any particular situation. It can include the decision not to use AI-DAPT system and/or one of the AI-DAPT tools in a specific case to establish levels of human discretion during the use of the same, or to ensure the ability to override a decision made by it.

- It should be ensured the human oversight and control over the decision cycles and operation of the AI-DAPT. It can be avoided only in exceptional cases, when demonstrating such oversight is not necessary (due to compelling reasons). Nevertheless, also in these cases it should be ensured (and explained) that humans will be able to understand the decisions made by the system and what mechanisms will exist for humans to override them.

- Specific training on how to exercise oversight should be given to the humans (human-in-the-loop, human-on-the-loop, human-in-command).

- Detection and response mechanisms should be established for undesirable adverse effects of AI-DAPT for the end-user or subject.

- A "stop button" or procedure to safely abort an operation when needed should be in place.

- Specific oversight and control measures should be taken to reflect the self-learning or autonomous nature of the AI-DAPT.

R3. Privacy and Data Governance

- In case the AI-DAPT system and/or an AI tool is trained or developed, by using or processing personal data (including special categories of personal data), the privacy and data protection implications of data collected, generated or processed over the course of the AI system's life cycle must be considered and such personal data must be collected and processed in a lawful, fair and transparent manner.

- The AI-DAPT system should be aligned with relevant standards (e.g. ISO25, IEEE26) or widely adopted protocols for (daily) data management and governance.

- The AI data governance models must include the Data Minimisation Principle and the Data-Protection-by-Design-and-by-Default paradigm (e.g. through measures like encryption, pseudonymisation, aggregation, anonymisation). As regards the Data Minimization Principle, it means that the personal data collected and or handled in AI-DAPT must be “adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed”. The anonymization and pseudonymization techniques should be adopted to the maximum extent, including safeguards for mitigating the risks of re-identifying the individuals and for minimizing possible linkability and actual linkages. As regards the Data-Protection-by-Design-and-by-Default paradigm, the AI-DAPT technical team should consider and gets aligned with the seven privacy principles defined by Cavoukian [159]: “1. Proactive not reactive – preventative not remedial 2. Privacy as the default setting 3. Privacy embedded into design 4. Full functionality – positive-sum, not zero-sum 5. End-to-end security – full lifecycle protection 6. Visibility and transparency – keep it open 7. Respect for user privacy – keep it individual and user-centric”.

- Adequate data governance mechanisms should be in place, covering the quality and integrity of the data used, their relevance in light of the domain in which the AI systems will be deployed, their access protocols and the capability to process data in a manner that protects privacy.

- The data should be acquired, stored, and used in a manner which can be audited by humans.

- Adequate technical and organisational measures for safeguarding the rights and freedoms of data subjects must be taken, such as the use of appropriate anonymization and/or pseudonymization techniques, the appointment and involvement of a data protection officer, encryption and aggregation methods.

- Strong security measures should be established for preventing data breaches and leakages, following the international security standards and ensuring compliance with the Cybersecurity Act.

- Mechanisms that allow flagging issues related to privacy concerning AI-DAPT and/or its AI tools should be in place, depending on the specific situation.

- The data processing must be performed according to the data protection legislation and any other applicable law and regulation, including those stemming from the national legislation (Lawfulness Principle). It encompasses the respect of the data protection obligations and requirements, described in the point below.

- The requirements and obligations set by the GDPR and other data protection legislations must be accomplished, including:
    1. The processing of personal data must rely on a legal basis. One of the legal basis is the informed consent, which is “any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear

affirmative action, signifies agreement to the processing of personal data relating to him or her" (art. 4 GDPR).

2. Purpose limitation and legitimate aim principle: the data must be collected for specific, explicit and legitimate purpose served by the AI-DAPT system/components, without further processing them in a way incompatible with it. Adequate safeguards against misuse must be taken.
3. The implementation of the Data Protection Impact Assessment (DPIA), when required by the GDPR or national legislation.
4. Appointment of a Data Protection Officer (DPO) and early involvement of him/her in the development or use phase of AI-DAPT.
5. Oversight mechanisms for data processing (including limiting access to qualified personnel, mechanisms for logging data access and making modifications).
6. Data Minimization Principle and Privacy-by-Design-and-by-Default approach (see above);
7. Storage Limitation Principle: the personal data must either be erased or anonymized as soon as they are no longer necessary for the purpose (Art. 5 (1) (e) GDPR);
8. The data subjects must be effectively entitled to exercise their rights, laid down in the Articles 12 –22 GDPR, including the right to withdraw consent, the right to object and the right to be forgotten into the development of the AI system. It is key to ensure the individual control of personal data, pursuant to both GDPR and the upcoming ePrivacy Regulation (ePR).

- The Data Portability must be ensured, according to both the Regulation on the free flow of non-personal data (art. 6) and the GDPR (art. 20), to enable or facilitate the switching of service providers and the porting of data between different IT systems, in a structured, commonly used and machine-readable format. The self-regulatory codes of conduct, respectively on data portability and on cloud switching, developed by the SWIPO Working Group - Switching from Provider and Porting non-personal data, should be taken into account.

R4. Technical Robustness, Security, Risk Assessment & Mitigation

- The Consortium must ensure the security, the safety and technical robustness of AI-DAPT. These requirements apply to both the data sharing and the AI-driven components/tools of the overall system, since its trustworthiness requires that it is able to deliver services that can justifiably be trusted (dependability), besides being robust when facing changes (resilience).

- The integrity, confidentiality and availability of the data must be ensured, including appropriate security of the data, especially the protection of the personal data against unauthorized or unlawful processing and against accidental loss, destruction or damage.

- Both for personal data processing (where relevant) and, in general, for AI design and development, appropriate technical and organizational measures should be implemented, considering the level of security appropriate to the risk (Art. 32 GDPR). Such measures should be taken, also to avoid cyber-security attacks (Art. 5, letter f GDPR).

- A risk-based approach should be adopted in AI-DAPT design, development, deployment and operation, pursuant to both the GDPR and by the AI Act. The former requires to evaluate the ethics risks related to the data processing activities by assessing the particular likelihood and severity of each risk to data protection, considering "the nature, scope, context and purposes of the processing and the sources of the risk". The risk assessment must be conducted in an objective manner to determine whether there is a "risk" or a "high risk". Obligations are set in case of high risks. Pursuant to Recital 75, 76 of GDPR, the risk level (in terms of likelihood and severity for freedoms and rights of individuals) determines what measures are appropriate in each case. The more severe and likely the risks are, the stronger measures will be required to counteract such risks. On the other hand, the AI Act classifies the AI system according to the risks posed by them and provides different obligations and requirements for the different types of systems.

- A preventative approach to risks should be adopted, during the development of AI-DAPT and/or its tools, in order to achieve technical robustness, reliable behaviour and minimizing unintentional and unexpected harm.
- An assessment of potential ethical risks associated with AI-DAPT system, including risk evaluation procedures and post-deployment mitigation measures, should be carried out. This risk assessment and mitigation exercise should be proportionate to the extent necessary for AI-DAPT research and expected future application.

- Authorization and Access Control mechanisms should be ensured. It is necessary to ensure that the participating users act according to the security, privacy and data sharing policies. Access to AI-DAPT technology and datasets should be possible only to authorized users.

- The data sovereignty must be guaranteed to incentivize the data sharing and build/reinforce trust among participants. In this direction, the IDS standard DIN SPEC 27070, should be taken into consideration. Metadata should be attached to data, unambiguously defining data usage policies at each level of the data value chain. The technical infrastructure should be able to enforce such data sovereignty, in order to make possible the execution of contractual provisions on the access and use of data, which, in turn, enforce the data policies in terms of processing, allow (or disallow) linkage or analysis of data-by-data users, or allow (or disallow) third parties access to data, and other use limitations, flow control, data transfer restrictions, etc. It might imply, in case of multiple parties involved, such as third parties' digital infrastructures (e.g. clouds, software components, networks), to proceed with Industry Agreements fostering data sharing schemes and practices.

R5. Fairness and Avoidance of Unfair Bias

- The Fairness Principle must be considered in the design and deployment of AI-DAPT technology, both in its substantive dimension, regarding an ideal of equal treatment between individuals or between groups of individuals, and in its procedural dimension, consisting in the ability to seek and obtain relief when individual rights and freedoms are violated. This fairness requirement relies especially on the Ethics Guidelines for Trustworthy AI, the GDPR (art. 5.1 a), as well as the Art. 21 European Charter of Fundamental Rights, which forbids any kind of discrimination, and on Declaration on European Digital Rights and Principles.

- The AI-DAPT Consortium must prevent that the AI systems suffer from the inclusion of inadvertent historic bias and incompleteness. It is functional to avoid that the overall solution and/or some of its components/tools facilitate any kind of discrimination or social sorting, or generate an undue or unjustified harm to anyone, including wrongfully stigmatization. Therefore, attention should be paid to avoid algorithmic bias, including the avoidance of bias in input data, modelling, and algorithm design. The data about people should be representative of the target population and reflect their diversity or be sufficiently neutral: the steps and measures to be taken to ensure this aspect should be clearly defined. Furthermore, it should be clarified how to identify and avoid bias in input data and in the algorithmic design, considering also the inferences drawn by the system, when it is capable of disadvantaging certain groups of people or single individuals.

- The Data Quality and Accuracy must be ensured the Consortium must ensure that the data are of high quality, accurate, consistent, and contextualized, taking every reasonable step to ensure to prevent the use of inaccurate data, in order to avoid that the data or AI-DAPT system or its components/tools lead to biased or erroneous outputs, untrustworthy results, lack of contextual relevance, and, ultimately, a loss of trust.

- Functional bias should be avoided by AI-DAPT system by offering the same level of functionality and benefits to end-users with different abilities, beliefs, preferences, and interests. The accessibility should be ensured: the universal accessibility principles should be adopted to the maximum possible extent, following the correspondent relevant accessibility guidelines, in order to ensure that the system is designed to be usable by different types of end-users with different abilities, regardless of their age, gender, abilities or other characteristics. It is also recommended that the User Interfaces (UI) are developed in a user and data protection friendly manner. Accessibility to the AI-DAPT system for persons with disabilities should be considered as well, referring to the Universal Design principles to address the widest possible array of users. Relevant accessibility standards should be followed.

- Besides avoiding the impacts resulting from algorithmic bias or lack of universal accessibility, the AI system should not negatively affect (in the short, medium and longer term) the interests of certain groups.

R6. Individual, Social and Environmental Well-being

- The AI-DAPT system must consider all end-users and stakeholders, avoiding to unduly or unfairly reduce their psychological and emotional well-being, but even contributing to their empower and advancing their interests and well-being.

- The AI-DAPT system itself and the supply chain to which it connects (or will connect in its operation in the post-project phase) should follow the principles of environmental sustainability at the maximum extent possible. The work to maximize the overall environmental impact of the AI-DAPT system should be documented. The solutions of the project should be based on choices aimed at minimizing the environmental impact and the carbon footprint. The sustainability and ecological responsibility of the overall AI-DAPT system and/or its tools should be pursued. The most environmentally friendly solutions should be selected in the system's development, deployment and use process, as well as its entire supply chain, examining the resource usage and energy consumption during training, opting for less harmful choices.

- The safety in the workplace (or, more in general, in the context of application of the AI-DAPT) must not be reduced by the AI-DAPT systems. The compliance with standards on workplace safety, employee integrity and compliance should be documented. The health and safety risks must be identified and prevented or minimized, adopting appropriate mitigating measures. The risks might be related to physical harm and comprise hazardous collisions, cybersecurity, lack of focus, loss of movement control, debris and pinch points, as well as they might consist in cognitive risks and psychological/ethical risks, including mental strain, lack of trust and complicated interaction mechanisms, social impact and acceptance. In the demonstrators of the project, all the relevant safety at work regulatory sources must be applied, the relevant health and safety procedures and protocols must be adopted and dedicated training must be organized to minimize the risks.

R7. Transparency, Traceability and Explainability

The Transparency requirement is set forth by different pieces of legislation, such as the GDPR, the AI Act, the RPLD Proposal and the AILD proposal. In particular:

- Must be made clear to the end-users and other stakeholders that they are interacting with an AI system and it must be communicated to them its purpose, capabilities, limitations, benefits, and risks (and those of the decisions conveyed by it), such as its level of accuracy and/ or error rates and including instructions on how to use the AI-DAPT system properly.

- The AI-DAPT system must be able to explain both its technical processes and the reasoning behind its decisions or predictions. The AI driven decisions – to the extent possible – must be explained to and understood by those directly and indirectly affected.

- The solutions should comprehensible, explainable or understandable from an external observer. The explainability requires that the AI-DAPT system is intelligible to non-experts, in particular those directly and indirectly affected. It occurs if its functionality and operations can be explained non-technically to a person not skilled in the specific domain. The degree to which explainability is needed depends on the context and the severity of the consequences of erroneous/ inaccurate output to human life.

- Measures must be put in place to enable the traceability of the AI-DAPT system during its entire lifecycle, i.e. the data and processes that yield its decisions must be properly documented. In other words, it must be made possible to track and document the journey of a data input and related processes through all stages of the data lifecycle within the processes of the development of the AI-DAPT system and its operation. Traceability comprises tracing back i) which data were used by the AI-DAPT system to make a certain decision(s) or recommendation(s), ii) which AI model or rules led to the decision(s) or recommendation(s) of the AI-DAPT system, as well as having adequate logging practices in place to record the decision(s) or recommendation(s) of the AI-DAPT system.

- Details about how decisions made by the AI-DAPT system can be explained to users must be provided, including the reasons behind the system's decisions. Explainability is key especially for systems that make decisions, recommendations, or take actions that could potentially cause significant harm or impact individual rights or interests.

- In case an explanation on why a model has generated a particular output or decision (and what combination of input factors contributed to that) is not possible (black-boxes), other explainability measures must be in pace (e.g. traceability, auditability and transparent communication on the AI system's capabilities).

- Records must be kept of all relevant decisions to allow tracing how ethical requirements (such as the removal of bias from a dataset) have been considered and met.

- It is recommended to follow the standards addressing transparency, such as the IEEE Standard for Transparency of Autonomous Systems (IEEE Std 7001TM-2021).

- It is recommended to provide the users with appropriate training material and disclaimers on how to adequately use the AI-DAPT system.

R8. Accountability-by-Design

The Accountability requirement, aimed at ensuring the evaluation of the AI-DAPT system by internal and external auditors and the access records on said evaluations, stems from different legal and ethical sources, including the GDPR, the AI Act, the AILD Proposal, the RPLD Proposal and the Ethics Guidelines for Trustworthy AI. It demands for the following aspects:

- Appropriate technical and organizational measures should be in place to be able to demonstrate responsibility and accountability for the AI-DAPT system operation and its outcomes (including the processing of data), and thereby the compliance with the legislations and ethical mandates.

- must be documented how possible undesirable effects of the AI-DAPT system will be detected, stopped, and prevented from reoccurring.

- The AI-DAPT system should be auditable by independent third parties, including both the decisions of the system itself and the procedures and tools used during the development process. The auditability of the AI-DAPT system must entail the enablement of the assessment of algorithms, data and design processes (providing verifiable evidence on the correctness of the current state of each component/system entities), and the existence of redress mechanisms in place in case of unjust adverse impacts.

- Mechanisms that facilitate the AI-DAPT system's auditability (e.g. traceability of the development process, the sourcing of training data and the logging of the AI system's processes, outcomes, positive and negative impact) must be in place.

- A risk management process should be in place to identify and tackle with potential vulnerabilities, risks or biases in the AI-DAPT system.

Compliance with the obligations and requirements set forth by the AI Act

The AI-DAPT system might be classified as a high-risk AI system or as a limited to minimal risk AI system, mainly depending on the context of application in the post-project phase.

Considering the specific circumstances of application in the future, the obligations and requirements respectively foreseen by the AI Act for the high-risk AI systems and for the limited to minimal risk AI systems must be respected.

The requirements provided by the AI Act for the **high-risk AI systems** regard various stages (from design and implementation to post-market introduction. These requirements for high-risk system, which will enter into force on 2 August 2026", set forth in Sect. 2 (art. 8-15 AI Act) are as follows:

- The Article 8 "Compliance with the Requirements" clarifies that the high-risk AI systems shall comply with these requirements, "considering their intended purpose as well as the generally acknowledged state of the art on AI and AI-related technologies" and that the providers shall be responsible for ensuring compliance.

- The Article 9 "Risk Management System" states that "a risk management system shall be established, implemented, documented and maintained in relation to high-risk AI systems". The risk management system shall be a "continuous iterative process planned and run throughout the entire lifecycle of a high-risk AI system, requiring regular systematic review and updating. It shall comprise" a set of pre-defined steps, listed in the provision itself. The risks relevant for this are those which "may be reasonably mitigated or eliminated through the development or design of the high-risk AI system, or the provision of adequate technical information». The risk management measures "shall be such that the relevant residual risk associated with each hazard, as well as the overall residual risk of the high-risk AI systems is judged to be acceptable". The "High-risk AI systems shall be tested for the purpose of identifying the most appropriate and targeted risk management measures. Testing shall ensure that high-risk AI systems perform consistently for their intended purpose and that they are in compliance with the requirements set out" by the AI Act. The "testing procedures may include testing in real-world conditions" and "shall be performed, as appropriate, at any time throughout the development process, and, in any event, prior to their being placed on the market or put into service. Testing shall be carried out against prior defined metrics and probabilistic thresholds that are appropriate to the intended purpose of the high-risk AI system".

- The Article 10 "Data and Data Governance" states that the "High-risk AI systems which make use of techniques involving the training of AI models with data shall be developed on the basis of training, validation and testing data sets that meet the quality criteria referred to in" such art. 10 whenever such data sets are used. Any training, validation and testing data sets "shall be subject to data governance and management practices appropriate for the intended purpose of the high-risk AI system", as well as "shall be relevant, sufficiently representative, and to the best extent possible, free of errors and complete in view of the intended purpose. They shall have the appropriate statistical properties, including, where applicable, as regards the persons or groups of persons in relation to whom the high-risk AI system is intended to be used. Those characteristics of the data sets may be met at the level of individual data sets or at the level of a combination thereof". The "data sets shall take into account, to the extent required by the intended purpose, the characteristics or elements that are particular to the specific geographical, contextual, behavioural or functional setting within which the high-risk AI system is intended to be used". Furthermore, provided that certain conditions are met and "to the extent that it is strictly necessary for the purpose of ensuring bias detection and correction in relation to the high-risk AI systems…", the providers "may exceptionally process special categories of personal data, subject to appropriate safeguards for the fundamental rights and freedoms of natural persons".

- The Article 11 "Technical Documentation" states that "the technical documentation of a high-risk AI system shall be drawn up before that system is placed on the market or put into service and shall be kept up-to date. The technical documentation shall be drawn up in such a way as to demonstrate that the high-risk AI system complies with" the requirements for high-risks systems and "to provide national competent authorities and notified bodies with the necessary information in a clear and comprehensive form to assess the compliance of the AI system with those requirements". Simplified technical documentation will be requested for small and microenterprises. In given circumstance, in case a high-risk AI system is placed on the market or put into service, a single set of technical documentation shall be drawn up.

- The Article 12 "Record-Keeping" asks that the high-risk AI systems "technically allow for the automatic recording of events (logs) over the lifetime of the system". "In order to ensure a level of traceability of the functioning of a high-risk AI system that is appropriate to the intended purpose of the system, logging capabilities shall enable the recording of events relevant for a set of given circumstances. The minimum aspects to be tracked by the logging capabilities are also listed and include, for instance, the "recording of the period of each use of the system (start date and time and end date and time of each use".

- The Article 13 "Transparency and Provision of Information to Deployers" states i) that the "high-risk AI systems shall be designed and developed in such a way as to ensure that their operation is sufficiently transparent to enable deployers to interpret a system's output and use it appropriately" and ii) that the high-risk AI systems "shall be accompanied by instructions for use in an appropriate digital format or otherwise that include concise, complete, correct and clear information that is relevant, accessible and comprehensible to deployers". The provision also comprises a list of information to be provided in such instructions.

- The Article 14 "Human Oversight" states that:
  1. The "high-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which they are in use.
  2. Human oversight shall aim to prevent or minimise the risks to health, safety or fundamental rights that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse, in particular where such risks persist despite the application of" the other requirements provided for high-risks systems.
  3. The oversight measures shall be commensurate with the risks, level of autonomy and context of use of the high-risk AI system, and shall be ensured through either one or both of the following types of measures:
     a. measures identified and built, when technically feasible, into the high-risk AI system by the provider before it is placed on the market or put into service.
     b. measures identified by the provider before placing the high-risk AI system on the market or putting it into service and that are appropriate to be implemented by the deployer.
  4. Tor implementing the oversight activities, "the high-risk AI system shall be provided to the deployer in such a way that natural persons to whom human oversight is assigned are enabled, as appropriate and proportionate:
     a. To properly understand the relevant capacities and limitations of the high-risk AI system and be able to duly monitor its operation, including in view of detecting and addressing anomalies, dysfunctions and unexpected performance;
     b. remain aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system (automation bias), in particular for high-risk AI systems used to provide information or recommendations for decisions to be taken by natural persons;
     c. To correctly interpret the high-risk AI system's output, considering, for example, the interpretation tools and methods available;
     d. To decide, in any particular situation, not to use the high-risk AI system or to otherwise disregard, override or reverse the output of the high-risk AI system;
     e. intervene in the operation of the high-risk AI system or interrupt the system through a 'stop' button or a similar procedure that allows the system to come to a halt in a safe state.

- The Article 15 "Accuracy, Robustness and Cybersecurity" states that the "high-risk AI systems shall be designed and developed in such a way that they achieve an appropriate level of accuracy, robustness, and cybersecurity, and that they perform consistently in those respects throughout their lifecycle". It will be encouraged the development of benchmarks and measurement methodologies to address the technical aspects of how to measure the appropriate levels of accuracy and robustness and "the levels of accuracy and the relevant accuracy metrics of high-risk AI systems shall be declared in the accompanying instructions of use". In addition, the "high-risk AI systems shall be as resilient as possible regarding errors, faults or inconsistencies that may occur within the system or the environment in which the system operates, in particular due to their interaction with natural persons or other systems. Technical and organisational measures shall be taken in this regard. The robustness of high-risk AI systems may be achieved through technical redundancy solutions, which may include backup or fail-safe plans. High-risk AI systems that continue to learn after being placed on the market or put into service shall be developed in such a way as to eliminate or reduce as far as possible the risk of possibly biased outputs influencing input for future operations (feedback loops), and as to ensure that any such feedback loops are duly addressed with appropriate mitigation measures". "High-risk AI systems shall be resilient against attempts by unauthorised third parties to alter their use, outputs or performance by exploiting system vulnerabilities". Technical solutions appropriate to the relevant circumstances and the risks will be taken.

Furthermore, pursuant to the Article 16 "Obligations of providers of the High-risk AI systems" of the AI Act (which will enter into force on 2 August 2026), the providers of high-risk AI systems must:

- Ensure that their high-risk AI systems are compliant with the requirements listed above.
- Indicate on the high-risk AI system or, where that is not possible, on its packaging or its accompanying documentation, as applicable, their name, registered trade name or registered trade mark, the address at which they can be contacted.
- Have a quality management system in place which complies with Article 17.
- Keep the documentation referred to in Article 18.
- When under their control, keep the logs automatically generated by their high-risk AI systems as referred to in Article 19.
- Ensure that the high-risk AI system undergoes the relevant conformity assessment procedure as referred to in Article 43, prior to its being placed on the market or put into service.
- Draw up an EU declaration of conformity in accordance with Article 47.
- Affix the CE marking to the high-risk AI system or, where that is not possible, on its packaging or its accompanying documentation, to indicate conformity with this Regulation, in accordance with Article 48.
- Comply with the registration obligations referred to in Article 49(1).
- Take the necessary corrective actions and provide information as required in Article 20.
- Upon a reasoned request of a national competent authority, demonstrate the conformity of the high-risk AI system with the requirements set out in Section 2.
- Ensure that the high-risk AI system complies with accessibility requirements in accordance with Directives (EU) 2016/2102 and (EU) 2019/882.

The Art. 40 "Harmonised Standards and Standardisation Deliverables" states that high-risk AI systems or general-purpose AI models that comply with certain standards will be considered in line with the requirements of the AI Act. On the other hand, pursuant to Art. 41 "Common Specifications", the EC can create common rules for AI systems if certain conditions (such as if the European standardisation organisations don't accept a request to create a standard, or if the standards created don't address fundamental rights concerns) are met and, if the AI systems meet these common rules, they are assumed to meet the requirements of the law.

The Art. 43 "Conformity Assessment" regards the demonstration of compliance of a high-risk AI system with the requirements set out for high-risks systems following a conformity assessment procedure in given circumstances.

In case of **limited to minimal risks**, the transparency requirement has to be followed, to ensure that humans are informed when necessary or that AI-generated content is identifiable as such.

The **General Purpose AI Models** (AI models can be used for many different purposes) and Generative AI must comply with specific transparency requirements, including: i) a declaration to users indicating that the content they are interacting with was generated by an AI system; ii) summaries of copyrighted data used in the training of these AI models must be provided. The more advanced AI models which might have a significant impact, such as GPT-4, are subject to an extensive evaluation and, in case of serious incidents these must be reported to the EC.

Responsible Use of Generative AI

According to the EC's Living guidelines on the Responsible use of Generative AI in research, which complements the provisions of the AI Act, the AI-DAPT technical team should respects its indications, as follows:

- The researchers, who are accountable for the integrity of the content generated by or with the support of generative AI, have to maintain a critical approach to using the output produced by generative AI, being of its limitations (such as potential bias, hallucinations and inaccuracies).
- The use of generative AI must be transparent, including detail on the specific generative AI model/tool used.
- It must be taken into account the stochastic (random) nature of generative AI tools, which might produce different output from the same input (to the detriment of reproducibility and robustness of the results and conclusions).
- Attention must be paid to privacy, confidentiality and intellectual property rights.

Considering that the generative AI tools are evolving quickly, it is key to stay up to date on the best practices.

8.3 Ethical principles, regulatory landscape and ethical and legal requirements for AI-DAPT Demonstrators

8.3.1 Demonstrator #1 – Health

8.3.1.1 Ethical Considerations, including HITL

The ethical considerations of the Demonstrator # 1 are described below in the table.

Ethics Implications	Description	Planned mitigating measures/ safeguards/ steps to be taken
Human-in-the-Loop (HITL)	Medical experts are involved in every step of the algorithm design to ensure that the data processed is physiologically relevant.	**Data collection**: The medical experts will perform the patient trials whilst adhering to professional medical standards. **Data reviewing**: Given the correct visualisation tools, these experts will also visually check that the data collected is coherent to medical knowledge. **Algorithm review**: The algorithm's performance will be presented to the medical experts. The predicted results should follow medically meaningful patterns. **Algorithm revision**: Based on their medical knowledge, instances of model performance degradation such as bias, drift and training-serving skew can be potentially linked to physiological patterns that can serve as a foundation for improving the algorithm through HITL.
Human-in-the-Loop (HITL)	Technical expertise and informed user cases.	As MCS makes its own hardware and software, the in-house experts are fundamental to fully understand and properly utilise the sensors and frameworks that collect the data. The AI-DAPT framework will provide the user with insight to the model's performance and suggestions to improve this performance. Thus, a human is introduced in the loop, and they can tweak the model as they see fit.
GDPR Compliance	HITL: Human oversight in data processing and collection. Ensure full adherence to GDPR standards.	Obtain informed consent, enforce robust data security measures, and manage data under the oversight of Charité's data office. Anonymise key identifiable information where possible.
Data Breach Prevention	HITL: Regular monitoring and control access.	Implement strict data protection protocols, restrict access to collected data, and control access to patient lists

	Prevent unauthorized access or misuse of participant data.	
Participant Bias	HITL: Human involvement in bias assessment. Recruitment limited to predominantly non-healthy senior Caucasian participants may affect generalizability.	Recognize this limitation to improve the generalisation of the glucose-measuring algorithm. Clearly specify its intended use case and label the algorithm to reflect any inherent bias. The training process of AI systems should be disclosed, and any biases or limitations should be addressed. Additionally, clear specification of tasks and conditions for responsible AI use should be given and human oversight should be insured. Algorithms' decision-making process/results should be explainable, that is, it should be possible to describe and understand how an algorithm arrives at a particular decision or outcome.
Misuse of Data	Data must only be used for algorithm development, not for medical device approval.	Securing the agreement of collaborating parties to use data solely for algorithm development purposes. HITL to monitor control access.
Algorithmic Bias	Risk of AI finding spurious, physiologically implausible correlations.	Implement a Human-in-the-loop (HITL) approach to validate data. Ensure that the AI system is explainable and clearly documented. Implementation of fairness metrics to ensure equitable outcomes.
Noise-Generating Factors	Physiological factors and excessively sensitive measurements can lead to noisy data, impacting glucose signal accuracy.	Conduct research on relevant factors to refine the algorithm, filter and clean the data based on a HITL approach.
Causality vs. Correlation	Machine learning may find correlations that aren't causative.	Evaluate factors for physiological relevance to improve model accuracy and reliability using a HITL approach.
Wearable Device Misuse	Human intervention and guidance in device usage. Incorrect device use may lead to inaccurate glucose predictions.	Use HITL approach to monitor and verify correct device usage.

Patient Data Privacy and Security in Data Processing	Ensure patient data collected during human interactions remains confidential and secure.	Use role-based access control and data encryption during transfer and storage along with anonymization, pseudonymization, and encrypted storage of the collected data.
Data Ownership/Data Sovereignty	Retain patient ownership and patient anonymity over the health data throughout the algorithm development process.	Pseudonymity allows the patients the option to opt that their data to be removed from the project. Human-supervised data sovereignty checks.
Transparency in Data Flow	If a system is not inspectable, the human-in-the-loop cannot take an informed decision about the data usage.	Develop interfaces that allow human auditors to check and control how the data is acquired, stored and used.

8.3.1.2 Ethical and Legal Framework

The legal and ethical reference sources applicable to Demonstrator #1 are described in the table below.

Regulatory source	Relevant content	Legal and/or ethical issues concerned	Other
Good clinical practice (ICH-GCP)	Provides international ethical and scientific quality standards for designing, conducting, recording, and reporting clinical trials involving human participants.	Ensures participant safety, data integrity, and quality in clinical trials. Relevant for maintaining ethical standards.	
Declaration of Helsinki (version of 2013)	Sets principles for medical research involving human subjects, emphasizing participant safety, informed consent, and risk-benefit considerations.	Establishes ethical framework for clinical research, including respect for participants and prioritizing their well-being over scientific or societal interests.	
EU General Data Protection Regulation (GDPR)	Establishes stringent data privacy and security requirements, including informed consent, data minimization, and rights to access, rectify, and delete personal data.	Protects individual privacy rights in handling personal data, essential for handling and securing participant data in compliance with EU regulations.	
Berlin Data Protection Act (BlnDSG)	Provides data protection requirements specific to the Berlin region,	Ensures adherence to regional data protection standards, important for	Aligned with GDPR but specific for Berlin.

	complementing GDPR with local guidelines.	managing participant data within Berlin's jurisdiction in line with both GDPR and local expectations.	
IEEE P7001 Standard for Transparency of Autonomous Systems	Provides guidelines on designing transparent AI systems, making the decision-making process understandable to end users.	Enhances trustworthiness of AI systems by promoting transparency in autonomous decision-making.	

8.3.1.3 Ethical and Legal Requirements

The legal and ethical requirements applicable to Demonstrator #1 are described in the table below.

Req #	EL Requirements	Description	Priority	Nature	AI-DAPT Technology Asset	Business Process/Objectives
01	ALTAI Requirement #1. Human agency and oversight	Ensure human autonomy in AI-driven decision-making processes. The user should have override options.	Critical	Ethical	Non-invasive glucose prediction algorithm.	Ensure user control and autonomy in health tasks.
02	ALTAI Requirement #2. Technical robustness and safety	Guarantee resilience and reliability in algorithmic functions and data analysis.	Critical	Ethical & Legal	Non-invasive glucose prediction algorithm.	Provide fail-safes or warnings when detecting unhealthy predictions to ensure safe use.
03	ALTAI Requirement #3. Privacy and Data Governance	Manage data securely, with stringent access control to protect patient information privacy.	Critical	Legal	Data Security framework.	Maintain privacy and integrity of the health data through encryption, anonymisation and access control.
04	ALTAI Requirement #4. Transparency	Ensure clear data lineage, interpretability, and documentation of AI decision processes.	Critical	Ethical	Data Observability, and HITL services.	Enhance transparency in data usage and patient care.

05	ALTAI Requirement #5. Diversity, non-discrimination and fairness	Monitor for biases, ensuring model fairness and inclusivity for diverse demographics.	Optional	Ethical	Non-invasive Glucose measurement algorithm validation process.	Highlight the bias within the available patient demographic. Attempt to reduce this bias through technologies such as synthetic data generation.
06	ALTAI Requirement #6. Societal & Environmental Well-being	Account for the societal impact, including environmental sustainability in AI systems	Optional	Ethical	Hardware requirements for Demonstrator 1: Health, solution.	Remain within existing wearable technological processing powers to avoid the need to replace existing wearables for this service to work through responsible AI practices.
07	ALTAI Requirement #7. Accountability	Define accountability measures for ethical compliance, risk management, and AI outcomes.	Critical	Ethical	Observability Module and Lineage tracking service.	Improve AI trustworthiness and accountability by ensuring accountability structures for AI decisions, with traceable logs for audits.
08	Ethical Proposal	The ethical proposal outlines the study's objectives, methodology, and potential risks and benefits.	Critical	Ethical and Legal	Non-invasive glucose prediction algorithm tested in a human clinical trial.	Ensures participant safety, data integrity, and quality in clinical trials.
09	Informed Consent	Participants provide informed consent through a written consent form that clearly explains the study's details and their rights.	Critical	Ethical and Legal	Non-invasive glucose prediction algorithm tested in a human clinical trial.	Ensures free will to participate in a clinical trial.

10	Participant Information Sheet	This document provides essential information about the study to potential participants.	Critical	Legal	Non-invasive glucose prediction algorithm tested in a human clinical trial.	Ensures participants receiving a comprehensive written explanation about the clinical trial. Participants are fully informed and understand what they are consenting to.
11	Data Sharing Plan	A clear plan outlining how data will be shared and used.	Critical	Legal	Non-invasive glucose prediction algorithm tested in a human clinical trial.	Protects individual privacy rights in handling and sharing personal data.
12	Authorization Concept	Procedures for authorizing access to data.	Critical	Legal	Non-invasive glucose prediction algorithm tested in a human clinical trial.	Protects individual privacy rights in handling and sharing personal data.
13	Data Protection Concept	Maintain strict data privacy by encrypting and anonymising sensitive health data before processing.	Critical	Legal	AI-DAPT pipeline storage systems and communication protocols.	Protects individual privacy rights in handling and sharing personal data.
14	Register of Processing Activities	A record of all data processing activities.	Critical	Legal	Non-invasive glucose prediction algorithm tested in a human clinical trial and lineage tracking for data flow within AI pipeline.	Protects individual privacy rights in handling and sharing personal data.
15	Risk Assessment	An evaluation of potential risks associated with data processing.	Critical	Legal	Data observability to ensure data health throughout the pipeline.	Aid debugging processes and provide transparency of the services happening.

16	Legal Contracts	Agreements governing data sharing with external parties.	Critical	Legal	Non-invasive glucose prediction algorithm tested in a human clinical trial.	Protects individual privacy rights in handling and sharing personal data.

8.3.2 Demonstrator #2 – Robotics

8.3.2.1 Ethical Considerations, including HITL

In Robotics & Cognitive Ergonomics demonstrator workers are equipped with wearable devices to monitor in real time the safety and stress condition of the operators, suggesting proper intervention. The demonstrator presents several legal and ethical considerations essential to its deployment. These considerations aim to protect workers' rights, uphold transparency, and ensure safe and ethical AI integration in the workplace. Ethical aspects include:

1. **Data and Algorithmic Bias**: The demonstrator relies on wearables collecting biometric and personal data to assess workers' fatigue and stress levels. However, data bias could occur if certain groups (e.g., gender, age, health state, lifestyle) are underrepresented in the collected data. To mitigate this, synthetic data will supplement real data to ensure diverse representation and minimize algorithmic bias.

2. **Transparency and Explainability**: Workers may not have full visibility into AI decision-making processes due to the black-box nature of some algorithms. While AI model technical explainability might be limited, efforts will be made to communicate the system’s capabilities and limitations clearly. Training will be provided to inform users about the data collection, its purpose and processing, and potential benefits.

3. **Privacy and Data Ownership**: Personal data gathered, including health and biometric indicators, need to be handled in compliance with GDPR standards. Strict data ownership protocols, data access limitations, and data minimization principles will be followed. Workers must provide informed consent, and the system will prioritize anonymization and data non-persistence.

4. **Human Dignity and Autonomy**: AI systems must respect human autonomy and avoid manipulative behaviours. Workers retain the right to remove wearable devices at will, preserving control over their participation. All collected data will be used solely to enhance safety, and not for surveillance or performance evaluations that could infringe upon their rights.

5. **Digital Divide**: The generational digital divide might impact users’ ability to interact with the wearable devices. To address this, the system will employ intuitive interaction notifications (e.g., vibrational alerts) and provide training to ensure all workers are comfortable using the technology.

6. **Safety and risk of harms, societal well-being**: The objective of the Demonstrator is to increase workers’ safety identifying stress and fatigue conditions for the operators. The alarm allows to monitor and intervene, supporting people’s physical and mental wellbeing. In this case, it does not represent an ethical issue, but the main objective of the demonstrator.

Human-in-the-loop (HITL) focus: Humans are the core of developed solutions and technology is at their service. The choice for the AI solution usage is completely up to the operator, wearing the devices or not according to the situation as well as the ability to override the anomaly detected by AI (human discretion).

- In the *training stage* of the developed solution, humans intervene to validate and regulate the decisions cycle of the system.
- In the *usage phase*, humans represent the source of data, and, at the end, they are informed about anomalies and could change task and/or prevent safety issues.

Ethics Implications	Description	Planned mitigating measures/safeguards/ steps to be taken
Data and Algorithmic Bias	Workers' data are collected through wearable devices and matched with personal data to monitor the fatigue of operators and prevent safety issues.	Synthetic data will supplement real data to ensure diverse representation and minimize algorithmic bias.
Transparency and Explainability	Presence of AI and the processes leading to AI system's decisions are not clearly visible during the usage.	Communicate the system's capabilities and limitations clearly. Training will be provided to inform users about the data collection, its purpose and processing, and potential benefits.
Privacy and Data Ownership	Legal and Ethical aspects related to Data.	Compliance with GDPR standards. Strict data ownership protocols, data access limitations, and data minimization principles will be followed. Workers must provide informed consent, and the system will prioritize anonymization and data non-persistence.
Human Dignity and Autonomy	Aspects related to the unproper use of the system to monitor and evaluate the performance of the workers.	Workers retain the right to remove wearable devices at will, preserving control over their participation. All collected data will be used solely to enhance safety, and not for surveillance or performance evaluations that could infringe upon their rights.
Digital Divide	Unequal access to digital technologies given the possible generational gap.	The system will employ intuitive interaction notifications (e.g., vibrational alerts) and provide training

		to ensure all workers are comfortable using the technology.
Human-in-the-loop	Human as part of the process.	In the *training stage* of the developed solution, humans intervene to validate and regulate the decisions cycle of the system. In the *usage phase*, humans represent the source of data, and, at the end, they are informed about anomalies and could change task and/or prevent safety issues.

8.3.2.2 Ethical and Legal Framework

The legal and ethical sources applicable to Demonstrator #2 are described in the table below.

Regulatory source	Relevant content	Legal and/or ethical issues concerned	Other
ISO/IEC 24745 Information security, cybersecurity and privacy protection — Biometric information protection	ISO/IEC 24745 covers the protection of biometric information under various requirements for confidentiality, integrity and renewability/ revocability during storage and transfer. It also provides requirements and recommendations for the secure and privacy-compliant management and processing of biometric information. This document does not include general management issues related to physical security, environmental security and key management for cryptographic techniques.	ISO/IEC 24745 provide the guidelines to be followed while handling biometrics data, in particular specifies: Analysis of the threats to and counter-measures inherent to biometrics and biometric system application models; Security requirements for securely binding between a biometric reference (BR) and an identity reference (IR); Biometric system application models with different scenarios for the storage and comparison of BRs; Guidance on the protection of an individual's privacy during the processing of biometric information.	
DIRECTIVE 2006/42/EC Of	It defines the essential safety and public health requirements	To be commercialized, the developed machines	

The European Parliament And Of The Council of 17 May 2006 on machinery	to which the machines must comply when they are designed, manufactured and operated before being placed on the market.	should be provided with a proper technical documentation and declaration of conformity.	
D.LGS. 81/2008 Testo Unico Salute e Sicurezza sul Lavoro National normative	The Act sets out the rights and obligations of employers, workers and other parties involved, defining the preventive measures to be taken to avoid accidents at work and occupational diseases. Promotes the culture of safety and health in the workplace, providing a series of rules and regulations to ensure the correct levels of safety and health of workers.	There are recommendations also for the evaluation of work-stress related risk. It also provides guidelines for the use of personal protective equipment and work equipment, including sensors, to ensure the safety of workers. In addition, companies that intend to install video surveillance systems at workplaces must obtain specific authorization.	
European Commission Ethics Guidelines for Trustworthy Artificial Intelligence	On 8 April 2019, the High-Level Expert Group on AI presented tasked by EC released a set of Guidelines that set out a framework for achieving Trustworthy AI.	The Guidelines put forward a set of 7 key requirements that AI systems should meet in order to be deemed trustworthy. A specific assessment list aims to help verify the application of each of the key requirements: 1. Human agency and oversight 2. Technical Robustness and safety 3. Privacy and data governance: a. Transparency b. Diversity, non-discrimination and fairness c. Societal and environmental well-being d. Accountability	Human agency and oversight, Non-discrimination and fairness, Societal and environmental well-being, Privacy and data governance checklist are particularly significant for this use case.
General Data Protection Regulation (GDPR)	The General Data Protection Regulation (GDPR) is a comprehensive data privacy law that establishes a framework for	To process data, it is necessary to consider seven protection and	

Regulation (EU) 2016/679	the collection, processing, storage, and transfer of personal data.	accountability principles outlined in Article 5.1-2: 1. Lawfulness, fairness and transparency — Processing must be lawful, fair, and transparent to the data subject. 2. Purpose limitation — Process data for the legitimate purposes specified explicitly to the data subject collected. 3. Data minimization — Collect and process only as much data as absolutely necessary for the purposes specified. 4. Accuracy — Keep personal data accurate and up to date. 5. Storage limitation — Store only personally identifying data for as long as necessary for the specified purpose. 6. Integrity and confidentiality — Processing must be done in such a way as to ensure appropriate security, integrity, and confidentiality (e.g. by using encryption). 7. Accountability — The data controller is responsible for being able to demonstrate GDPR compliance with all of these principles.	

8.3.2.3 Ethical and Legal Requirements

The legal and ethical requirements applicable to Demonstrator #3 are described in the table below.

Req #	EL Requirements	Description	Priority	Nature	AI-DAPT Technology Asset	Business Process/Objectives
01	Human agency and oversight	Ensures workers can override AI recommendations, retaining autonomy over task adjustments.	Critical	Ethical	Sensors (hardware)	Risk mitigation, Operator wellbeing.
02	Technical robustness and safety	Focuses on safe, reliable AI responses to mitigate risks like fatigue and stress detection errors and unproper sampling.	Critical	Ethical	Data acquisition / Data storage	Risk mitigation, Operator wellbeing, reliable solution, representativeness of different conditions.
03	Privacy and data governance	GDPR compliance for personal data handling, including data minimization and worker consent.	Critical	Legal	Data collection / Data Storage / Data lifecycle management	Representativeness of different conditions, anonymization.
04	Transparency	Ensures workers understand the AI's capabilities and limitations.	Preferred	Ethical	\	Risk mitigation, Operator wellbeing, Operator autonomy.
05	Diversity, non-discrimination	Prevents bias by integrating synthetic data for demographic representativeness.	Preferred	Ethical	Synthetic Data Generation Engine	Risk mitigation, Operator wellbeing, Workers' inclusion.
06	Societal and environmental well-being	AI aims to enhance workers' physical and mental well-being, increasing safety and ergonomics.	Preferred	Ethical	\	Risk mitigation, Operator wellbeing.
07	Accountability	Establishes protocols for data access and ethical use to maintain transparency in data handling.	Critical	Legal	Data Storage, Pipeline Management	Risk mitigation, Operator wellbeing.
08	Biometrics Data Handling	Protection of biometric information under various requirements for confidentiality, integrity and	Critical	Legal	Data acquisition, Data Storage, Data lifecycle	Risk mitigation, Operator wellbeing.

		renewability/revocability during storage and transfer.			management	
09	Stakeholders' involvement	Workers will be involved on voluntary basis to demonstrate the developed solution. Need of informative consent.	Critical	Ethical	Data lifecycle management, sensors	Increase safety and operator wellbeing.
10	Ethics Guidelines for Trustworthy Artificial Intelligence (EC)	The Guidelines put forward a set of 7 key requirements that AI systems should meet in order to be deemed trustworthy.	Preferred	Ethical	\	Risk mitigation, Operator wellbeing.

8.3.3 Demonstrator #3 – Energy

8.3.3.1 Ethical Considerations, including HITL

This section gives examples and highlights the importance of ethical practices in Pilot 3 (Energy), especially in safeguarding personal data of end-users. Zenith and Domx, as pilot collaborators and joint data controllers, are responsible for ensuring data subjects' (end-users') rights, privacy and freedoms and autonomy are protected. This ethical commitment aligns with both the GDPR and relevant national (Greek) legislation, ensuring the lawful and transparent processing of personal data. By obtaining prior consent from individuals (end-users) and maintaining confidentiality, the parties (Zenith and Domx) demonstrate a commitment to ethical data management and respect for individual privacy and autonomy.

Description of Risks

Key risks associated with Pilot 3 (Energy) involve potential data breaches, unauthorised access and improper data handling. Such incidents could lead to the unlawful disclosure or manipulation of personal data, compromising the privacy and security of data subjects (end-users). Given that data processing includes sensitive information, such as heating system performance and personal identifiers, risks also encompass potential unauthorised profiling and automated decision-making.

Planned Mitigating Measures

To mitigate such potential risks, the partners (Zenith and Domx) have outlined several technical and organisational security measures. These include:

- Controlled Access: Restricting access to data to authorised personnel only.
- Use of Secure Environments: Protecting equipment with firewalls and storing it in secure, controlled-access locations.
- Monitoring and Regular Updates: Continuous monitoring of equipment and software, along with daily backups to prevent data loss.
- Data Breach Protocols: Promptly notifying the other party and documenting any data breaches to enable swift corrective actions.

- Employee Training: Conducting regular training to ensure staff are aware of and compliant with data protection protocols.

The table below provides more information on the issue at hand.

Ethics Implications	Description	Planned mitigating measures/ safeguards/ steps to be taken
GDPR Compliance	Both parties (Zenith and Domx) must align with GDPR regulations, upholding the legal rights of data subjects (end-users) and adhere to high standards of data protection and privacy. There should be established clear responsibilities, commitment to lawful processing, transparency, and fairness, protecting data subjects' privacy and freedoms. Failure to comply with GDPR requirements could lead to significant legal and financial risk. Risks arise from improper data processing, lack of adequate consent management, or failure to uphold individuals' rights, such as data deletion or access rights.	Zenith and Domx will ensure robust GDPR compliance by enforcing consent protocols, allowing data subjects to access, modify or delete their data as required. Data Protection Officers (DPOs) from both parties will oversee data compliance, conduct audits and manage privacy rights requests.
Data and Algorithmic Bias	An ethical consideration is ensuring that data processing and any algorithms applied in the pilot do not introduce bias, especially in energy efficiency interventions. Bias could affect outcomes, impacting certain users unfairly. Ethical responsibility includes regular evaluation of algorithms to detect and correct biases to ensure fair treatment. Bias risks in algorithmic assessments, particularly concerning energy efficiency or heating system performance, may lead to inaccurate conclusions that unfairly disadvantage certain participants. This can undermine the validity of the pilot and result in biased or unethical outcomes.	Both parties (Zenith and Domx) will perform ongoing evaluation and adjustments to algorithms to prevent bias, ensuring that data assessments remain fair. Both ZeniΘ and Domx will regularly review and validate algorithmic models to detect biases and enhance fairness.
Transparency and Explainability	Transparency is ensured by requiring that data subjects (end-users) are informed of how their data will be used, allowing them to understand and consent to the process. This commitment to transparency aims to prevent mistrust and misunderstandings about data use and decision-making processes, thus reinforcing ethical accountability. If data flows, processing purposes or algorithms are not adequately explained, data subjects may feel misled or suspicious, resulting in ethical and legal complaints. Lack of transparency can damage	Transparent data practices are essential. Data subjects (end-users) will receive clear explanations on how their data will be used and the logic behind any automated decision-making. This includes providing easily accessible information on data flows and enabling subjects to withdraw their consent at any time.

	trust and potentially lead to misinformed consent.	
Privacy and Data Ownership	Respecting data ownership rights of subjects (end-users) is central to the pilot's ethical approach. Zenith and Domx recognise data subjects' sovereignty over their data, emphasising that data will only be used with explicit consent and as intended under the pilot agreement. Individuals retain rights to access, modify or delete their data as per GDPR guidelines. A risk exists that the data subjects' ownership rights may be compromised if data is accessed or processed without adequate consent. Inadequate handling or unauthorised use of data may lead to loss of trust and legal challenges.	To uphold data ownership rights, both parties (Zenith and Domx) will draft a detailed agreement including thorough consent mechanisms and limits data use strictly to the agreed-upon purposes. Subjects will retain control over their data and have access to it upon request, ensuring respect for data sovereignty.
Comfort and Autonomy	Pilot parties (Zenith and Domx) prioritise protecting the comfort and autonomy of data subjects (end-users), ensuring that data processing does not infringe upon their comfort or autonomy.	Any intervention intended to assess energy efficiency and demand-side management, will prioritise participants' comfort and autonomy by incorporating explicit consent and enabling withdrawal at any time.
Human-in-the-Loop	Algorithmic outcomes will incorporate human oversight. To ensure ethical safeguards (e.g. managing the heating system of the end-user, the set-point temperature, flow temperature, operational hours, etc.), human oversight is maintained within data processing activities.	Human review of automated decisions will allow for ethical judgment and error correction, ensuring that automated assessments do not negatively impact data subjects' rights (e.g., there will be a feedback mechanism from the end-user perspective, as well as end-user overwrite authority of any automated heating management decision).
Data Breach	Data breaches could expose sensitive information, leading to privacy violations and financial, legal and reputational damage. Risks include unauthorised access by external parties, malicious attacks or accidental internal breaches.	Zenith and Domx will draft a detailed agreement which will outline stringent security protocols, including: **Technical Controls**: Use of firewalls, encryption and controlled access measures to protect data from unauthorised access. **Organizational Controls**: Training employees on data

		security best practices, restricting access based on roles, and enforcing confidentiality agreements. **Data Breach Protocols**: In the event of a breach, immediate notification to affected parties and supervisory authorities, alongside containment and corrective measures.
Misuse of Data	Improper use of collected data for purposes beyond the pilot's objectives or unauthorised access by third parties, risks violating data subjects' privacy and GDPR guidelines.	To prevent data misuse, both parties commit to limiting data access strictly to necessary personnel and preventing data from being repurposed without new consent. Data access logs and audits will monitor and prevent unauthorised use.
Transparency in Data Flow	If data flows are not transparent, participants may misunderstand how data is shared or processed across entities, leading to distrust and potential GDPR violations if data is accessed in ways subjects did not explicitly consent to.	To maintain transparency, a record of data flows and processing activities will be made available to data subjects. Regular audits and data flow diagrams will provide additional transparency, ensuring data subjects can trace how their data is processed and shared.

8.3.3.2 Ethical and Legal Framework

The legal and ethical sources applicable to Demonstrator #3 are described in the table below.

Regulatory source	Relevant content	Legal and/or ethical issues concerned	Other
General Data Protection Regulation (EU) 2016/679 27 April 2016 Regulation (EU) 2016/679 Binding EU Regulation	Article 5 outlines core principles, such as lawful, fair, and transparent processing, purpose limitation, and data minimisation. Article 6 specifies lawful bases for processing, such as consent and contract performance. Article 25 requires data protection by design and	The GDPR is foundational for data protection in the EU and is directly relevant to the future joint controllership agreement between Zenith and Domx for the pilot implementation. It mandates clear responsibilities and compliance standards, ensuring lawful data processing,	The GDPR requires comprehensive documentation and transparent processing practices. Non-compliance may lead to fines, and full compliance enhances trust with participants by ensuring high data privacy standards.

	by default, encouraging companies to embed privacy features from the outset. Article 32 mandates implementing security measures to protect data. Articles 33 and 34 cover data breach notifications, requiring prompt reporting to authorities and affected individuals. Articles 35-36 mandate Data Protection Impact Assessments (DPIAs) for high-risk data processing activities.	transparency, security and accountability. Since the pilot involves processing customer data, GDPR compliance underpins all data-related activities, aligning the agreement with EU legal standards.	
Greek Law 4624/2019 29 August 2019 4624/2019 National legislation implementing GDPR requirements	Law 4624/2019 elaborates GDPR applications for data handling within Greece (Energy Pilot), including provisions for supervisory authorities, data subject rights and lawful processing requirements. It outlines data processing rules for both the public and private sectors, adapting GDPR's broader standards to specific national contexts.	This law directly applies to the Pilot (Zenith-Domx) agreement as it further specifies GDPR requirements within Greece. Adherence is essential to uphold data subjects' rights and aligns the agreement with the Greek Data Protection Authority's standards, which apply when processing Greek customers' data.	Law 4624/2019 emphasises the importance of collaborating with Greece's supervisory authority for data processing guidance, adding an extra layer of accountability for Zenith and Domx.
Ethics Guidelines for Trustworthy AI April 2019 Non-binding EU guidelines	These guidelines set principles for ethical AI use, emphasising transparency, accountability, fairness and human oversight. They recommend avoiding algorithmic bias, maintaining user autonomy and ensuring AI systems operate within ethical boundaries.	Since the pilot may involve automated decision-making in energy management, adherence to these guidelines can prevent ethical concerns, such as bias or lack of transparency. This adds an ethical layer to the data-sharing framework, ensuring fairness and respect for user autonomy.	Though non-binding, these guidelines align with the EU's AI regulatory framework, reflecting best practices for responsible AI development and deployment.

8.3.3.3 Ethical and Legal Requirements

The legal and ethical requirements applicable to Demonstrator #3 are described in the table below.

Req #	EL Requirements	Description	Priority	Nature	AI-DAPT Technology Asset	Business Process/Objectives
01	ALTAI Requirement #1. Human agency and oversight	Human-in-the-Loop for Automated Decisions: Ensuring human oversight in AI-driven calculations to allow for intervention and review in automated decisions that impact data subjects (end-users).	Critical	Ethical	Energy management solutions with human review capabilities for automated processes, enabling intervention when necessary.	AI-based decision-making processes, especially in offering energy efficiency and personalised energy management.
02	ALTAI Requirement #1. Human agency and oversight	Data Subject Consent Management: Obtaining and managing informed consent from data subjects to ensure their agency over how their data is collected, processed and shared.	Critical	Legal	Consent management mechanism that capture, document and track consent records, ensuring transparency and user control.	Consent acquisition and management processes, affecting all stages of energy management.
03	ALTAI Requirement #2. Technical robustness and safety	Data Breach Prevention and Incident Response: Establishing robust security measures to prevent data breaches and setting protocols for effective incident response.	Critical	Legal	Firewalls, encryption, breach detection tools and incident response processes.	Security management processes across data storage, access and transmission.
04	ALTAI Requirement #2. Technical	Data Integrity and Secure Processing:	Critical	Legal	Encrypted/Anonymised storage, data integrity monitoring tools and	Data storage and processing procedures,

	robustness and safety	Implementing measures to ensure data integrity, secure processing environments and resilient data access controls.			secure server environments.	especially for data analytics and energy assessments.
05	ALTAI Requirement #3. Privacy and data governance	Data Protection by Design and by Default: Integrating privacy and data protection measures at every stage of data processing, from initial collection to deletion.	Critical	Legal	Privacy-enhancing technologies, encryption, access management systems and secure data handling protocols.	Data lifecycle management, including collection, storage, processing, and disposal stages.
06	ALTAI Requirement #3. Privacy and data governance	Confidentiality and Non-Disclosure of Personal Data: Ensuring data confidentiality, restricting disclosure and preventing unauthorized access in compliance with GDPR.	Critical	Legal	Encryption tools, secure data storage, confidentiality agreement and controlled data access systems.	Internal data handling and employee data access protocols.
07	ALTAI Requirement #3. Privacy and data governance	Data Ownership and Sovereignty Assurance: Respecting data subjects' rights over their data, ensuring data sovereignty and complying with their data control preferences.	Critical	Legal	Data storage and sharing processes, emphasizing data subjects' control over their information.	Consent tracking tools and access control platforms for data ownership management.
08	ALTAI Requirement #4 Transparency	Transparency and Accountability in Data Processing: Ensuring data subjects receive clear information	Critical	Ethical & Legal	User-accessible privacy dashboards, automated transparency tools, and detailed data flow diagrams.	Customer communication, privacy policy disclosures, and consent processes.

		on how their data is collected, processed and used, in compliance with GDPR and ethical transparency standards.				
09	ALTAI Requirement #4 Transparency	Transparent Data Flow and Audit Trail: Maintaining clear documentation of data flow and audit trails for accountability and traceability in data handling.	Preferred	Ethical & Legal	Data flow monitoring tools, audit log management software and data tracking systems.	Data processing and record-keeping, particularly in data transfers between Zenith and Domx.
10	ALTAI Requirement #4 Transparency	Protection Against Algorithmic Bias: Preventing discrimination and bias in AI algorithms to ensure fair treatment of all data subjects in automated energy assessments, recommendations and demand-side management interventions.	Preferred	Ethical	AI and data analytics processes, especially in personalised energy recommendations and demand-side management interventions.	Bias detection and mitigation algorithms, fairness audit tools and bias assessment frameworks.
11	ALTAI Requirement #4 Transparency	Data Minimization and Purpose Limitation: Collecting only the necessary data for specific purposes, as outlined by GDPR, to avoid overreach and unnecessary processing of personal data.	Critical	Legal	Data filtering and anonymisation technique and purpose-limitation controls in data collection systems.	Data collection and processing procedures, especially in restricting data sharing between Zenith and Domx.

12	ALTAI Requirement #4 Transparency	Sustainable Data Usage: Prioritising efficient data usage practices in alignment with the environmental goals of the energy sector, ensuring energy efficiency in digital operations.	Preferred	Ethical & Legal	Data usage monitoring and optimisation tools.	Data processing and analytics, focusing on optimising energy use and reducing environmental impact.
13	ALTAI Requirement #5 Accountability	Data Access Limitation and Role-Based Access Control: Restricting data access to only authorised personnel based on job roles and enforcing accountability in data handling practices.	Critical	Legal	Role-based access control (RBAC) systems, identity management tools and employee access logging systems.	Internal data access management and processing within Zenith and Domx.
14	ALTAI Requirement #5 Accountability	Audit and Documentation of Data Processing Activities: Maintaining detailed documentation and records of all data processing activities, ensuring traceability and compliance with GDPR's accountability principles.	Critical	Legal	Documentation management software, audit trail systems and GDPR-compliant record-keeping tools.	Compliance and internal auditing processes across all stages of data processing.

8.3.4 Demonstrator #4 – Manufacturing

8.3.4.1 Ethical Considerations, including HITL

In demonstrator #4 for AI-driven maintenance processes for hoisting equipment in the aviation sector, we address key legal and ethical considerations to ensure responsible AI integration. Our approach prioritizes worker privacy, skill enhancement, fairness, compliance and sustainability. By implementing pseudonymization, anonymization, and obtaining active consent, we avoid intrusive surveillance and power imbalances. Rather than replacing workers, AI complements the daily skills set, enhancing their roles in maintenance tasks. Sensitive decisions remain human-in-the-loop, ensuring human judgment in critical scenarios.

To avoid prejudiced outcomes, we mitigate biases in data, fostering fairer decision-making. Intellectual property rights (IPR) are respected throughout the AI pipeline, upholding ethical use of proprietary information. Regulatory compliance and adherence to ethical standards guide our approach, ensuring the AI implementation aligns with aviation industry norms. Lastly, we use AI-DAPT tools to enable sustainable AI practices, minimizing environmental impact. These measures collectively support an ethical, human-centred, and compliant AI deployment in aviation maintenance.

Ethics Implications	Description	Planned mitigating measures/safeguards/ steps to be taken
Avoid Worker Surveillance and Imbalance of Power	Risk of employee discomfort and mistrust due to perceived surveillance. Potential power imbalances if management uses AI to closely monitor workers' productivity without consent. Legal and ethical risks associated with inadequate data privacy protections.	**Data Anonymization and Pseudonymization**: Implement strong data anonymization and pseudonymization practices to protect personal information. **Active Consent Mechanisms**: Obtain explicit consent from employees before collecting or using any data that could be seen as intrusive. **Transparent Communication**: Regularly communicate with employees about what data is being collected, how it will be used, and the benefits of AI for their work processes to foster trust and transparency.
Complement, Don't Replace, Existing Skill Sets	Fear of job displacement or deskilling among workers if AI is seen as a threat to their roles. Potential resistance from workers and unions if AI is introduced without considering workers' existing skills and roles.	**Training and Upskilling Programs**: Implement training programs that upskill workers to work alongside AI, enhancing their existing skill sets. **Clear Role Definition**: Define clear roles where AI complements rather than replaces human tasks to reinforce that AI is a tool to assist, not replace, workers. **Employee Involvement:** Engage employees in the AI deployment process, gathering feedback to ensure AI applications are aligned with their workflow and skills.
Human-in-the-Loop (HITL)	Over-reliance on AI in critical decisions could reduce human judgment, leading to errors in high-stakes scenarios. Risk of workers being disengaged or unprepared to intervene effectively if they are not regularly involved in decision-making processes.	**Mandatory Human Oversight**: Ensure that a human is required to review and approve all AI-driven decisions in sensitive areas, such as safety or compliance checks. **Regular Training and Scenario Testing**: Conduct regular training on scenarios where human intervention may be necessary, so that workers are prepared to override AI when needed. **Clear Escalation Paths**: Establish protocols for quickly escalating issues to a human operator in

		cases where AI makes an uncertain or anomalous recommendation. ***Examples** are: (i) the documentation of processes and their proper execution. AI will support in a gain of data quality, where AI might tend to supply with unhealthy drifts in data quality. Audits with the process stakeholders will be done for regular validation. (ii) inappropriate tool handling will be shown by AI-driven data analysis. Possible suspects (groups not individuals) related to unhealthy will be educated and process responsibilities will be invoked.*
Avoid Biases in the Data	Risk of biased decision-making if the data used to train AI models reflects historical or systemic biases. Potential for unfair treatment of certain groups if AI models amplify biases present in data.	**Diverse and Representative Data Collection:** Ensure data used for training AI systems is diverse and representative of different scenarios, tasks, and populations. **Bias Audits and Regular Monitoring**: Regularly audit and monitor AI models for biases and adjust them as needed to ensure fair and equitable outcomes. **Algorithmic Fairness Techniques**: Use algorithmic techniques that reduce or correct bias, such as re-sampling, re-weighting, or de-biasing methods in model training.
Respect Intellectual Property Rights (IPR)	Potential legal issues if AI models infringe on intellectual property rights by using proprietary or sensitive data without proper permissions. Risk of data misuse or IP theft, leading to loss of competitive advantage or legal penalties.	**Clear IP Policy and Agreements**: Develop clear IP policies and agreements for data usage, ensuring all parties understand their rights and responsibilities. **Access Controls**: Implement strict access controls and usage restrictions on proprietary data to prevent unauthorized use. **Regular Compliance Reviews**: Conduct regular compliance reviews to ensure that AI models and data usage respect IP rights and adhere to legal agreements.
Ensure Compliance with Regulations and Ethical Standards	Risk of non-compliance with industry regulations, leading to fines, reputational damage, or operational disruption. Ethical concerns if AI systems are deployed in ways that conflict with social or regulatory standards.	**Regulatory Compliance Checks:** Establish a checklist of regulatory requirements specific to aviation and AI applications to ensure compliance. **Ethics Committee or Review Board**: Set up an ethics committee or board to review AI applications from an ethical standpoint before deployment. **Regular Training on Regulations**: Train all stakeholders, including developers, operators,

		and managers, on relevant regulations and ethical standards.
Sustainable Implementation of AI (AI-DAPT Tools)	High energy consumption of AI models could increase the environmental footprint, clashing with sustainability goals. Risk of resource strain if AI infrastructure is not designed for efficiency and scalability.	**Energy-Efficient Models and Hardware**: Choose energy-efficient AI models and hardware that reduce power consumption and environmental impact. **Optimize Resource Usage**: Use cloud resources effectively and recycle computing resources where possible to reduce waste. **Continuous Sustainability Audits**: Conduct periodic audits to evaluate the sustainability impact of AI implementations and adjust improve resource efficiency and reduce carbon footprint.

8.3.4.2 Ethical and Legal Framework

The framework of the demonstrator is set by technical standards, considering aspects of interoperability, data collection and exchange, system integration and architectures, which are overlapping with ethical and legal aspects. The identified standards are listed and described in the following:

Regulatory source	Relevant content	Legal and/or ethical issues concerned	Other
ISO 13374 – Condition Monitoring and Diagnostics of Machines	This standard provides guidelines for data processing, communication, and presentation in condition monitoring and diagnostics.	It supports the development of systems that continuously monitor equipment health, which is critical for predictive maintenance in aviation.	n.a.
ISO 14224 – Collection and Exchange of Reliability and Maintenance Data for Equipment	This standard defines the requirements for gathering and sharing reliability and maintenance data.	ISO14224 helps in identifying and analysing equipment failures, enhancing the accuracy of maintenance processes.	n.a.
ISA-95 – Enterprise-Control System Integration	ISA-95 offers a framework for integrating enterprise and control systems, bridging the gap between operational technology (OT) and information technology (IT).	For the demonstrator, this facilitates seamless communication between AI-driven maintenance tools and enterprise systems.	n.a.
MIMOSA OSA-EAI – Open System Architecture for Enterprise Application Integration	MIMOSA OSA-EAI defines standards for interoperability among asset management systems.	This is essential for enabling data exchange and interoperability between various maintenance, engineering, and business systems within the demonstrator.	n.a.

IEC 62541 – OPC Unified Architecture (OPC UA)	This standard establishes a communication protocol for industrial automation systems, allowing secure and reliable data exchange.	OPC UA supports the integration of AI and IoT in maintenance systems by providing a unified communication framework.	n.a.
ISO 50000 – Energy Management Systems	ISO 50000 series focuses on energy management and efficiency, guiding organizations in setting up systems to reduce energy consumption.	For the demonstrator, this standard helps ensure that AI-driven processes align with sustainability and energy efficiency goals.	n.a.

8.3.4.3 Ethical and Legal Requirements

The following table outlines the ethical and legal (EL) requirements essential for implementing an AI-enhanced demonstrator in maintenance processes of hoisting equipment. These requirements address critical aspects such as data privacy, fairness, transparency, worker rights, regulatory compliance, intellectual property, sustainability, and ongoing ethical oversight. This structured approach ensures that the AI demonstrator adheres to ethical standards, legal mandates, and sustainable business practices, fostering trust and accountability in its deployment.

Req#	EL Requirements	Description	Priority	Nature	AI-DAPT Technology Asset	Business Process/ Objectives
1	ALTAI Requirement #3. Privacy and data governance: “Data Privacy and Security”	Anonymize and pseudonymize personal data; obtain consent and comply with GDPR	Critical	Ethical	Data Design for AI: Data Documentation following Data Sculpting/Nurturing/Curation for AI: Data selection and cleaning	Data Handling, Compliance
2	ALTAI Requirement #5. Diversity, non-discrimination and fairness: Fair and Unbiased AI	Conduct audits to prevent biases and ensure diverse data	Preferred	Ethical	Data Design for AI: Data Valuation	Decision-Making, Quality Control
3	ALTAI Requirement #4 Transparency & ALTAI Requirement #7. Accountability: Transparency	Communicate AI capabilities and maintain transparency in decisions	Preferred	Ethical	Data Sculpting/Nurturing/Curation for AI: Data Annotation	Trust Building, Risk Management

	and Accountability					
4	ALTAI Requirement #6. Societal and environmental well-being: Worker Rights and Empowerment	Avoid surveillance, protect worker autonomy, and involve workers in AI development	Critical	Legal	Data Design for AI: Data Documentation	Workforce Management, Employee Engagement
5	Compliance with Aviation / contractor and Safety Regulations	Adhere to safety standards and ensure human oversight in critical decisions	Critical	Legal	Data Design for AI: Data Documentation	Safety Compliance, Quality Assurance
6	ALTAI Requirement #7. Accountability: Respect for Intellectual Property Rights (IPR)	Protect proprietary data, establish IP agreements, and comply with IP laws	Preferred	Ethical	Data Design for AI: Data Documentation	IP Management, Competitive Advantage
7	ALTAI Requirement #6. Societal and environmental well-being: Environmental Sustainability	Use energy-efficient AI and minimize environmental impact	Preferred	Ethical/Legal	Model delivery for AI: Model evaluation	Sustainability, Resource Efficiency
8	ALTAI Requirement #1. Human agency and oversight: Continuous Compliance and Ethical Reviews	Perform regular assessments (set up ethics guide lines)	Optional	Ethical	Data Design for AI: Data Valuation	Governance, Continuous Improvement

8.4 Ethics and Data Protection Impact Assessment Methodology

AI-DAPT Trustworthy Framework also comprises the Ethics and Data Protection Impact Assessment (EDPIA) Methodology for the demonstrators, whose core aspects are sketched out in this paragraph, whilst its complete elaboration will be described in WP5 “Demonstration, Benchmarking, Business Validation and Impact Assessment”, and in particular in D5.1 - Demonstrators Evaluation Framework and Use Case Plan (M15). The EDPIA is directed to assess the risks for individuals' rights, freedoms and wellbeing, as well as for ensuring compliance with the data protection law (GDPR and national regimes), and ethical mandates for the research with humans, the protection of personal data and the design and/or use of Artificial Intelligence solutions, taking into account the data lifecycle and use cases in each of the AI-DAPT Pilots. The EDPIA will be conducted before the involvement of human beings and the collection of their personal data will take place and its findings will be reported in D5.2 Demonstrators' Activities Implementation Results – Draft Version (M30) and updated or refined at the end of the project in D5.3 - Demonstrators' Activities Implementation Results – Final Version (M42). This tool supports accountability and allows the demonstrator partner to show their respective due diligence in taking adequate actions to ensure full compliance on an ongoing basis in each use case scenarios, when relevant. The EDPIA will rotate around three components:

- Humans, which regards the participation in the piloting operations of individuals, volunteers and stakeholders external to the research staff
- Personal Data, focusing on the privacy, regulatory compliance and ethical implications regarding the collection and/or processing of personal data in the pilot concerned
- Artificial Intelligence, concerning the check on the applicability of the Ethics guidelines for trustworthy AI developed by the High-Level Expert Group on AI, the classification of the AI system to be used in the pilot on the basis of the AI Act risk-level approach, as well as insights and considerations from a human-rights perspective.

The table to be used to conduct the Ethics and Data Protection Impact Assessment for each of the AI-DAPT Pilots, gathering the pilot-specific features from the legal and ethical point of view, will be inserted in D5.1.

9 Conclusions and Next Steps

This deliverable serves as an output for tasks T1.1 - Automated AI Pipelines End-User Needs, Ethics Analysis, Constraints and Considerations, T1.2 - Automated AI Pipeline End-to-End Usage Scenarios and Hybrid Science-Guided AI Models Foundations, T1.3 - AI-DAPT Research & Technology Radar, T1.4 - Automated AI Pipeline Lifecycle Design & AI-DAPT Research Agenda, and T1.5 - AI-DAPT User Stories, Technical & AI Requirements under WP1 - Automated AI Pipeline Lifecycle Management Framework.

D1.2 has outlined the critical groundwork for the AI-DAPT framework, starting with an updated research agenda in Section 2, which focuses on key advancements in AI technologies and methods. This agenda will be used as the foundation for the iterative development of technical work packages and highlight the core research topics that AI-DAPT will address.

Section 3 provided a thorough overview of the Demonstrator pilots, emphasizing their data sources, updated user needs, and the identification of open datasets that will support platform development.

Section 4 introduced a preliminary design for the data and AI pipeline lifecycle, offering a clear step-by-step framework tailored to the envisioned processes of the pilots. This initial design sets the foundation for flexible, scalable integration across AI-DAPT applications.

Section 5 detailed the end-to-end usage scenarios, aligning workflows with user needs and technical requirements to ensure stakeholder expectations are met effectively.

Section 6 and 7 captured the functional and non-functional requirements of the platform, informing the iterative development process of the Minimum Viable Product (MVP).

Finally, Section 8 explored the ethical and legal dimensions of AI-DAPT, providing a robust framework for ensuring compliance, trustworthiness, and responsible deployment. This included tailored ethical considerations for each Demonstrator, addressing key aspects such as human-in-the-loop integration, data protection, and adherence to regulatory frameworks.

In conclusion, this deliverable establishes the foundation for AI-DAPT's technical and ethical objectives, paving the way for subsequent development in WP1 and beyond. It ensures that the framework is well-positioned to address the needs of its Demonstrators while meeting the highest standards of innovation, compliance, and usability. Moving forward, the work presented in D1.2 will continue to evolve through iterative updates, aligning with the broader goals of the AI-DAPT project.

10 References

[1] Y. Roh, G. Heo and S. E. Whang, “A Survey on Data Collection for Machine Learning: A Big Data-AI Integration Perspective,” *IEEE Transactions on Knowledge and Data Engineering,* pp. 1328-1347, 2019.

[2] M. Jovic, E. Tijan, R. Marx and B. Gebhard, “Big Data Management in Maritime Transport,” *Pomorski Zbornik,* pp. 123-141, 2019.

[3] “Apache Spark,” [Online]. Available: https://spark.apache.org/.

[4] “Apache Flink,” [Online]. Available: https://flink.apache.org/.

[5] “Amazon EMR (Elastic Map-Reduce),” [Online]. Available: https://aws.amazon.com/emr/.

[6] “What is Apache Flink? - Architecture,” [Online]. Available: https://flink.apache.org/what-is-flink/flink-architecture/. [Accessed 11 November 2024].

[7] “What is Amazon EMR,” [Online]. Available: https://docs.aws.amazon.com/pt_br/emr/latest/ManagementGuide/emr-what-is-emr.html. [Accessed 11 November 2024].

[8] “Apache Kafka,” [Online]. Available: https://kafka.apache.org/.

[9] “Amazon Kinesis,” [Online]. Available: https://aws.amazon.com/kinesis/.

[10] “RabbitMQ,” [Online]. Available: https://www.rabbitmq.com/. [Accessed 11 November 2024].

[11] “Introduction,” [Online]. Available: https://kafka.apache.org/intro. [Accessed 11 November 2024].

[12] “Apache Druid,” [Online]. Available: https://druid.apache.org/.

[13] “MinIO AIStor,” [Online]. Available: https://min.io/product/aistor-overview. [Accessed 14 November 2024].

[14] “Introduction to Apache Druid,” [Online]. Available: https://druid.apache.org/docs/latest/design/. [Accessed 14 November 2024].

[15] J. Conde, A. Pozo, A. Munoz-Arcentales, J. Choque and Á. Alonso, “Fostering the integration of European Open Data into Spaces through High-Quality Metadata,” *arXiv:2402.06693,* 2024.

[16] N. Thalhath, M. Nagamori and T. Sakaguchi, “Metadata application profile as a mechanism for semantic interoperability in FAIR and open data publishing,” *Data and Information Management,* p. 100068, 2024.

[17] “Data Catalog Vocabulary (DCAT),” [Online]. Available: https://www.w3.org/TR/vocab-dcat-3/.

[18] “Linking data: Data Catalogue Vocabulary Application Profile,” 22 November 2022. [Online]. Available: https://data.europa.eu/en/publications/datastories/linking-data-data-catalogue-vocabulary-application-profile. [Accessed 14 November 2024].

[19] "DCAT Application Profiles (DCAT-AP)," [Online]. Available: https://ec.europa.eu/isa2/solutions/dcat-application-profile-data-portals-europe_en/.

[20] K. Jiang, W. Liang, J. Y. Zou and Y. Kwon, "Opendataval: a unified benchmark for data valuation," *Advances in Neural Information Processing systems,* vol. 36, 2023.

[21] R. H. L. Sim, X. Xu and B. K. H. Low, "Data Valuation in Machine Learning: "Ingredients", Strategies, and Open Challenges," in *IJCAI*, 2022.

[22] M. Huang and R. Rust, "A strategic framework for artificial intelligence in marketing," *Journal of the Academy of Marketing Science,* vol. 49, pp. 30-50, 2021.

[23] "AI Fairness 360 - IBM," [Online]. Available: https://aif360.res.ibm.com/.

[24] Z. Tan, D. Li, S. Wang, A. Beigi, B. Jiang, A. Bhattacharjee, M. Karami, J. Li , L. Cheng and H. Liu, "Large Language Models for Data Annotation: A Survey," *arXiv:2402.13446,* 2024.

[25] M. Sutharsan, "Smart analysis of automated and semi-automated approaches approaches to data annotation for machine learning," *ICTACT Journal on Data Science and Machine Learning,* 2023.

[26] F. Demrozi, C. Turetta, F. a. Machot, G. Pravadelli and P. H. Kindt, "A Comprehensive Review of Automated Data Annotation Techniques in Human Activity Recognition," *arXiv:2307.05988,* 2023.

[27] W. X. K. Z. J. L. T. T. X. W. Y. H. Y. M. e. a. Zhao, "A survey of large language models," in *arXiv*, 2023.

[28] B. a. M. v. d. S. van Breugel, ""Position: Why Tabular Foundation Models Should Be a Research Priority," in *Forty-first International Conference on Machine Learning (ICML)*, 2024.

[29] X. Fang, W. Xu, F. Tan, J. Zhang, Z. Hu, Y. Qi, S. Nickleach, D. Socolinsky, S. Sengamedu and C. Faloutsos, " Large language models (LLMs) on tabular data: Prediction, generation, and understanding-a survey," *Transactions on Machine Learning Research,* 2024.

[30] X. Zhang, R. Chowdhury, R. Gupta and J. Shang, "Large language models for time series: A survey," *arXiv preprint,* 2024.

[31] P. Li, Y. He, D. Yashar, W. Cui, S. Ge, H. Zhang, D. Rifinski Fainman, D. Zhang and S. Chaudhuri, "Table-GPT: Table Fine-tuned GPT for Diverse Table Tasks," *Proceedings of the ACM on Management of Data,* vol. 2, no. 3, pp. 1-28, 2024.

[32] M. Eltabakh, Z. Naeem, M. Ahmad, M. Ouzzani and N. Tang, "RetClean: Retrieval-Based Tabular Data Cleaning Using LLMs and Data Lakes," *Proceedings of VLDB Endow.,* pp. 4421-4424, 2024.

[33] T. Zhou, P. Niu, L. Sun and R. Jin, "One fits all: Power general time series analysis by pretrained lm," *Advances in neural information processing systems,* vol. 36, pp. 43322-43355, 2023.

[34] G. Kumar, S. Basri, A. A. Imam and A. O. Balogun, "Data Harmonization for Heterogeneous Datasets in Big Data - A Conceptual Model," in *Software Engineering Perspectives in Intelligent Systems: Proceedings of 4th Computational Methods in Systems and Software 2020*, 2020.

[35] J. A. Patel and P. Sharma, “Big Data Harmonization - Challenges and Applications,” *IJRITCC,* pp. 206-208, 2017.

[36] A. L. Fusco Giuseppe, “An approach for semantic integration of heterogeneous data sources,” *PeerJ Cmputer Science,* p. e254, 2020.

[37] B. Abdelghani, “Data reconciliationand fusion methods: A survey,” *applied computing and Informatics,* pp. 182-194, 2022.

[38] H. L. Z. J. H. J. Z. K. Wang Feng, “A semantics-based approach to multi-source heterogeneous information fusion on the internet of things,” *Soft Computing,* pp. 2005-2013, 2017.

[39] D. A. V. G. M. M. M. Bianchini Devis, “Smart city Data modelling using semantic wen technologies,” in *2021 IEEE International Smart Cities Conference (ISC2)*, Manchester, United Kingdom, 2021.

[40] C. R. G. G. B. L. S. N. J.-G. R. Figueiras Paulo, “Big Data Harmonization for Intelligent Mobility: A Dynamic Toll-Charging Scenario,” in *On the Move to Meaningful Internet Systems: OTM 2016 Workshops: Confederated International Workshops: EI2N, FBM, ICSP, Meta4eS, and OTMA 2016*, Rhodes, Greece, 2017.

[41] VesselAI, “D2.3 VesselAI Extreme-scale data processing, management services and semantics,” Ch. 2.1-2.3. [Online]. Available: https://cordis.europa.eu/project/id/957237/results.

[42] N. Patki, R. Wedge and K. Veeramachaneni, “The synthetic data vault,” in *IEEE international conference on data science and advanced analytics (DSAA)*, 2016.

[43] N. Chawla, K. Bowyer, L. Hall and W. Kegelmeyer, “SMOTE: synthetic minority over-sampling technique,” *Journal of artificial intelligence research, 16,* pp. 321-357, 2002.

[44] H. Han, W. Wang and B. Mao, “Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning,” in *International conference on intelligent computing*, Berlin, 2005.

[45] C. Bunkhumpornpat, K. Sinapiromsaran and C. Lursinsap, “Safe-level-smote: Safe-level-synthetic minority over-sampling technique for handling the class imbalanced problem,” in *Advances in Knowledge Discovery and Data Mining: 13th Pacific-Asia Conference, PAKDD 2009*, Bangkok, 2009.

[46] H. He, Y. Bai, E. A. Garcia and a. S. Li, “ADASYN: Adaptive synthetic sampling approach for imbalanced learning,” in *IEEE international joint conference on neural networks*, Hong Kong, 2008.

[47] H. Murtaza, A. M. N. Khan, G. Z. S. Murtaza and A. Bano, “Synthetic data generation: State of the art in health care domain,” in *Computer Science Review*, 2023.

[48] W. J, K. M, N. J, Q. A, M. C, H. D, D. C, D. K, G. T and M. S., “Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record,” in *Journal of the American Medical Informatics Association*, 2018.

[49] Y. Liu, R. Stouffs and Y. Theng, “Development of Synthetic Patient Data to Support Urban Planning for Public Health,” in *ECAADe*, 2020.

[50] D. Kaur, M. Sobiesk, S. Patil, J. Liu, P. Bhagat, A. Gupta and N. Markuzon, “Application of Bayesian networks to generate synthetic health data,” *Journal of the American Medical Informatics Association,* vol. 28, no. 4, 2021.

[51] M. Baowaly, C. Lin, C. Liu and K. Chen, “Synthesizing electronic health records using improved generative adversarial networks,” *Journal of the American Medical Informatics Associatio,* vol. 26, no. 3, 2019.

[52] T. B. Brown , B. Mann , N. Ryder , M. Subbiah , J. Kaplan , P. Dhariwal , A. Neelakantan , P. Shyam , G. Sastry , A. Askell , S. Agarwal , A. Herbert-Voss, G. Krueger , T. Henighan , R. Child , A. Ramesh , D. M. Ziegler , J. Wu , C. Winter , C. Hesse , M. Chen , E. Sigler , M. Litwin , S. Gray , B. Chess , J. Clark , C. Berner , S. McCandlish , A. Radford , I. Sutskever and D. Amodei, “Language Models are Few-Shot Learners,” *arXiv preprint arXiv:2005.14165,* 2020.

[53] R. d. Lemos and M. Grzes, “Self-Adaptive Artificial Intelligence,” in *IEEE/ACM 14th International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS)*, Montreal, 2019.

[54] R. Bommasani, D. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, M. Bernstein, J. Bohg, A. Bosselut, E. Brunskill and E. Brynjolfsson, “On the opportunities and risks of foundation models,” in *arXiv preprint*, 2021.

[55] X. Han, Z. Zhang, N. Ding, Y. Gu, X. Liu, Y. Huo, J. Qiu, Y. Yao, A. Zhang, L. Zhang and W. Han, “Pre-trained models: Past, present and future,” in *AI Open*, 2021.

[56] M. T. Ribeiro, S. Singh and 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*, 2016.

[57] M. Christoph, Interpretable machine learning: A guide for making black box models explainable, 2020.

[58] C. Molnar and T. Freiesleben, Supervised Machine Learning for Science: How to stop worrying and love your black box, 2024.

[59] H. Baniecki, G. Casalicchio, B. Bischl and P. Biecek, “On the Robustness of Global Feature Effect Explanations,” *Joint European Conference on Machine Learning and Knowledge Discovery in Databases,* pp. 125-142, 2024.

[60] J. Herbinger, M. Wright, T. Nagler, B. Bischl and G. Casalicchio, “Decomposing global feature effects based on feature interactions,” *arXiv preprint,* 2023.

[61] K. D. M. H. S. C. H. S. H. K. K. P. L. J. M. S. M. A. M. S. N. K. N. R. J. T. R. D. S. Rachel K. E. Bellamy, “AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias.,” 2018.

[62] Gerards, Janneke and Xenidis, Raphaele , “Algorithmic Discrimination in Europe: Challenges and Opportunities for Gender Equality and Non-Discrimination Law,” European Commission, 2021.

[63] “European Union,” [Online]. Available: https://www.europarl.europa.eu/charter/pdf/text_en.pdf.

[64] S. M. H. a. A. N. Barocas, Fairness and machine learning: Limitations and opportunities, MIT Press, 2023.

[65] J. R. Sahil Verma, Fairness definitions explained. In Proceedings of the International Workshop on Software Fairness, New York, NY, USA: Association for Computing Machinery, 2018.

[66] J. H. a. E. N. Arjun Roy, "Multi-dimensional discrimination in law and machine learning - a comparative overview," in *ACM Conference on Fairness, Accountability, and Transparency, FAccT '23*, 2023.

[67] S. N. A. R. a. Z. S. W. Michael Kearns, " Preventing fairness gerrymandering: Auditing and learning for subgroup fairness," in *International Conference on Machine Learning*, 2017.

[68] A. B. Shiflet and G. W. Shiflet, Introduction to computational science: modeling and simulation for the sciences, Princeton University Press, 2014.

[69] E. Winsberg, Science in the age of computer simulation, University of Chicago Press, 2019.

[70] K. Cranmer, J. Brehmer and G. Louppe, "The frontier of simulation-based inference," *Proceedings of the National Academy of Sciences,* vol. 117, no. 48, pp. 30055-30062, 2020.

[71] N. E. R. D. J. M. S. L. B. Papamakarios George, "Normalizing Flows for Probabilistic Modeling and Inference," *arXiv preprint arXiv:1912.02762,* 2019.

[72] J. M. Lueckmann, J. Boelts, D. Greenberg, P. Goncalves and J. Macke, "Benchmarking simulation-based inference," in *International conference on artificial intelligence and statistics*, 2021.

[73] D. M. N. J. M. Greenberg, "Automatic posterior transformation for likelihood-free inference," in *International Conference on Machine Learning*, 2019.

[74] Z. Y. C. Y. L. Wang, "Development of RC model for thermal dynamic analysis of buildings through model structure simplification," *Energy and Buildings,* vol. 195, pp. 51-67, 2019.

[75] J. J. A. I. C. F. D. B. K. A. D. K. E. P. O. e. a. Drgoňa, "All you need to know about model predictive control for buildings," *Annual Reviews in Control,* vol. 50, pp. 190-232, 2020.

[76] A. Nagabandi, I. L. S. Clavera, R. Fearing, P. Abbeel, S. Levine and C. Finn, "Learning to adapt in dynamic, real-world environments through meta-reinforcement learning," *arXiv preprint,* 2018.

[77] F. Lafto and F. Afghah, "Meta-Reinforcement Learning Approach for Adaptive Resource Optimization in O-RAN," *arXiv preprint,* 2024.

[78] L. Zintgraf, S. Schulze, C. Lu, L. Feng, M. Igl, K. Shiarlis, Y. Gal, K. Hofmann and S. Whiteson, "Varibad: Variational bayes-adaptive deep rl via meta-learning," *Journal of Machine Leanring Research,* vol. 22, no. 289, 2021.

[79] V. Gudepu, V. R. Chintapalli, P. Castoldi, L. Valcarenghi, B. R. Tamma and K. Kondepu, "Adaptive retraining of ai/ml model for beyond 5g networks: A predictive approach," in *IEEE 9th International Conference on Network Softwarization (NetSoft)*, 2023.

[80] V. Gudepu, B. Chirumamilla, V. Chintapalli, P. Castoldi, L. Valcarenghi, B. Tamma and K. Kondepu, "Generative-AI for AI/ML Model Adaptive Retraining in Beyond 5G Networks," in *arXiv preprint*, 2024.

[81] Y. C. Z. L. G. a. E. M. Liang, "A new, short-recorded photoplethysmogram dataset for blood pressure monitoring in China," in *Scientific data*, 2018, pp. 1-7.

[82] S. S. A. S. A. a. T. R. Gupta, “DSVRI: A PPG-based novel feature for early diagnosis of type-II diabetes mellitus,” *IEEE sensors letters,* pp. 1-4, 2022.

[83] “WESAD: Multimodal Dataset for Wearable Stress and Affect Detection,” [Online]. Available: https://ubicomp.eti.uni-siegen.de/home/datasets/icmi18/.

[84] “A Wearable Exam Stress Dataset for Predicting Cognitive Performance in Real-World Settings,” [Online]. Available: https://www.physionet.org/content/wearable-exam-stress/1.0.0/.

[85] “Simultaneous physiological measurements with five devices at different cognitive and physical loads,” [Online]. Available: https://www.physionet.org/content/simultaneous-measurements/1.0.2/.

[86] “Multilevel Monitoring of Activity and Sleep in Healthy People,” [Online]. Available: https://www.physionet.org/content/mmash/1.0.0/.

[87] “EEG During Mental Arithmetic Tasks,” [Online]. Available: https://www.physionet.org/content/eegmat/1.0.0/.

[88] “Electrocardiogram, skin conductance and respiration from spider-fearful individuals watching spider video clips,” [Online]. Available: https://www.physionet.org/content/ecg-spider-clip/1.0.0/.

[89] “flEECe, an Energy Use and Occupant Behavior Dataset for Net Zero Energy Affordable Senior Residential Buildings (for 9 months),” [Online]. Available: https://osf.io/2ax9d/.

[90] “Municipal Building Energy Usage (2009-2014),” [Online]. Available: https://data.wprdc.org/dataset/municipal-building-energy-usage.

[91] “I-BLEND, a campus scale commercial and residential buildings electrical energy dataset,” [Online]. Available: https://springernature.figshare.com/collections/I-BLEND_a_campus_scale_commercial_and_residential_buildings_electrical_energy_dataset/3893581/1.

[92] “Automated Contract Review with Natural Language Inference,” [Online]. Available: https://stanfordnlp.github.io/contract-nli/.

[93] “Pile of Law,” [Online]. Available: https://huggingface.co/datasets/pile-of-law/pile-of-law.

[94] “Contract Summarization,” [Online]. Available: https://github.com/lauramanor/legal_summarization.

[95] “Pandas,” [Online]. Available: https://pandas.pydata.org/.

[96] “Pandas Dataframe Correlation Method,” [Online]. Available: https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.corr.html.

[97] “Scikit Learn PCA Method,” [Online]. Available: https://scikit-learn.org/dev/modules/generated/sklearn.decomposition.PCA.html.

[98] “UMAP: Uniform Manifold Approximation and Projection,” [Online]. Available: https://github.com/lmcinnes/umap.

[99] "Scikit Learn Linear Discriminant Analysis Method," [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.discriminant_analysis.LinearDiscriminantAnalysis.html.

[100] "Matplotlib," [Online]. Available: https://matplotlib.org/stable/users/index.html.

[101] "Seaborn," [Online]. Available: https://seaborn.pydata.org/tutorial.html#.

[102] "Scikit learn Feature selection method," [Online]. Available: https://scikit-learn.org/stable/modules/feature_selection.html.

[103] "PyWavelets," [Online]. Available: https://pywavelets.readthedocs.io/en/latest/ref/index.html.

[104] "Statsmodels: Statistics and Tests methods," [Online]. Available: https://www.statsmodels.org/stable/api.html#statistics-and-tests.

[105] "SciPy: Discrete Fourier transforms methods," [Online]. Available: https://docs.scipy.org/doc/scipy/reference/fft.html.

[106] "PyEMD EMD method," [Online]. Available: https://pyemd.readthedocs.io/en/latest/emd.html.

[107] "PyWavelets: Wavelet Transforms methods," [Online]. Available: https://pywavelets.readthedocs.io/en/latest/.

[108] "pys: A Python Package for Time Series Classification," [Online]. Available: https://pyts.readthedocs.io/en/stable/index.html.

[109] "NeuroKit2: Neurophysiological Data Analysis," [Online]. Available: https://neuropsychology.github.io/NeuroKit/.

[110] "MNE: Open-source Python package for exploring, visualizing, and analyzing human neurophysiological data," [Online]. Available: https://mne.tools/stable/index.html.

[111] "Time-series Generative Adversarial Networks (TimeGAN)," [Online]. Available: https://github.com/jsyoon0823/TimeGAN.

[112] "TimeGAN: A pytorch implementation of Time-series Generative Adversarial Networks," [Online]. Available: https://github.com/benearnthof/TimeGAN.

[113] "Nixtla: Open Source Time Series Ecosystem," [Online]. Available: https://github.com/Nixtla/.

[114] T. K. L. P. H. P. G. Zhu, "GluGAN: Generating Personalized Glucose," *IEEE Journal of Biomedical and Health Informatics,* vol. 10, no. 27, pp. 5122-5133, 2023.

[115] "PyTorch," [Online]. Available: https://pytorch.org/.

[116] "Scikit Learn," [Online]. Available: https://scikit-learn.org/stable/.

[117] "Darts: Time Series Made Easy in Python," [Online]. Available: https://unit8co.github.io/darts/.

[118] "TensorFlow Keras," [Online]. Available: https://www.tensorflow.org/guide/keras.

[119] "Hugging Face," [Online]. Available: https://huggingface.co/docs/transformers/index.

[120] "Keras Optimizers," [Online]. Available: https://keras.io/api/optimizers/.

[121] "LightGBM Python Package," [Online]. Available: https://lightgbm.readthedocs.io/en/latest/Python-Intro.html.

[122] "Scikit Learn Metrics," [Online]. Available: https://scikit-learn.org/stable/api/sklearn.metrics.html.

[123] "Keras Metrics," [Online]. Available: https://keras.io/api/metrics/.

[124] "PyTorch Metrics," [Online]. Available: https://pytorch.org/docs/stable/index.html.

[125] [Online]. Available: https://github.com/mlflow/mlflow.

[126] P. Achimugu, A. Selamat, R. Ibrahim and M. Mahrin, "A systematic literature review of software requirements prioritization research," *Information and software technology,* vol. 56, no. 6, pp. 568-585, 2014.

[127] H. AI, "High-level expert group on artificial intelligence," *Ethics guidelines for trustworthy AI,* vol. 6, 2019.

[128] P. Ala-Pietil, Y. Bonnet, U. Bergmann, M. Bielikova, C. Bonefeld-Dahl, W. Bauer, L. Bouarfa, R. Chatila, M. Coeckelbergh, Dignum and V. a. others, The assessment list for trustworthy artificial intelligence (ALTAI), European Commission, 2020.

[129] B. Dainow and P. Brey, "Ethics by design and ethics of use approaches for artificial intelligence," *European Commission DG Research \& Innovation RTD,* 2021.

[130] [Online]. Available: https://www.ai-dapt.eu/ .

[131] H. Graux, "What is data ownership, and does it still matter under EU data law? An exploration of traditional concepts of data ownership, and of the expected impact of the Data Act," *"Data Ownership and Access to Data - Position Statement of the Max Planck Institute for Innovation and Competition of 16 August 2016 on the Current European Debate,* 2016.

[132] S. Hao, W. Han, T. Jiang, Y. Li, H. Wu, C. Zhong, Z. Zhou and H. Tang, "Synthetic data in AI: Challenges, applications, and ethical implications," *arXiv preprint,* 2024.

[133] E. D. P. Supervisor, "TechSonar," *2021-2022 Report,* 2021.

[134] M. Heder, "AI and the resurrection of Technological Determinism," *Információs Társadalom: Társadalomtudományi Folyóirat,* 2021.

[135] W. S, "Technological determinism: what it is and why it matters," 2023.

[136] [Online]. Available: https://www.dataedgeusa.com/.

[137] M. M. F. e. al., "Towards Trustworthy AI: A Review of Ethical and Robust Large Language Models," *arXiv preprint,* 2024.

[138] Y. e. a. Wang, "AI Alignment: A Comprehensive Survey," *arXiv preprint,* 2023.

[139] E. Commission, "Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics," *COM,* vol. 64, 2020.

[140] E. P. Committee, “European Parliament resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence (2020/2014,” *INL,* 2020.

[141] E. Commission, “White Paper on Artificial Intelligence - A European Approach to Excellence and Trust,” *COM,* 2020.

[142] European Parliament, “Resolution on a civil liability regine for artificial intelligence, (2020/2014 (INL)),” *Resolution on a framework of ethical aspects of artificial intelligence, robotics and related technologies (2020/2012 (INL),* 2020.

[143] “Regulation (EU) 2024/1689”.

[144] “Proposal for a Directive of the European Parliament and of the Council on liability for defective product,” vol. COM 495 final, 2020.

[145] J. Cobbe, M. S. A. Lee and J. Singh, “Reviewable automated decision-making: a framework for accountable algorithmic systems,” 2021.

[146] Linkedin, “Human in the Loop: the Key to Achieving Explanability in AI,” 2024.

[147] W. Xingjiao and e. al., “A survey on Human-in-the-Loop for Machine Learning,” 2022.

[148] O. Gomez-Carmona and e. al., “Human-in-the-loop machine learning: Reconceptualizing the role of the user in interactive approaches,” *Internet of Things,* vol. 25, p. 101048, 2024.

[149] frontiere.io, “Can there be harmony between Human and AI? The key role of Explanable AI and Human in the Loop,” 2024.

[150] M. Chromik, M. Eiband and e. al., “I think I get your point, AI! The illusion of explanatory depth in explainable AI,” 2021.

[151] AI-DAPT, “D1.1 AI-DAPT Automated AI Pipeline sEnd User Needs and Scientific Technology Radar,” 2024.

[152] European Union , “European Union Agency for Fundamental Rights, Bias in Algorithms, Artificial Intelligence and Discrimation,” 2022.

[153] [Online]. Available: https://artificialintelligenceact.eu .

[154] European Union, “Communication on boosting startups and innovation in trustworthy artificial intelligence,” 2024.

[155] Shaping Europe's digital future, “Ethics guidelines for trustworthy AI,” 2019.

[156] European Research Area Forum, “Living guidelines on the responsible use of Generative AI in research,” 2024.

[157] European Union, “Guidance on the Regulation on a framework for the free flow of non-personal data in the European Union,” 2019.

[158] European Commision, “Ethics by Design and Ethics of Use Approaches for Artificial Intelligence,” 2021.

[159] A. Cavoukian, “Privacy by Design. The 7 Foundational Principles,” 2011.

[160] [Online]. Available: https://pycaret.org/.

[161] [Online]. Available: https://imbalanced-learn.org/stable/.

[162] F. Z. Z. G. W. A. K. a. G. C. Fahimi, "Towards EEG generation using GANs for BCI applications," in *IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, Chicago, 2019.

[163] V. G. J. a. K. A. Sushko, "One-shot gan: Learning to generate samples from single images and videos," in *IEEE/CVF conference on computer vision and pattern recognition*, Virtually, 2021.

[164] [Online]. Available: https://github.com/zhangzc21/Generalized-One-shot-GAN-Adaptation.

[165] [Online]. Available: https://airflow.apache.org/.

[166] [Online]. Available: https://docs.readthedocs.io/en/stable/.

[167] [Online]. Available: https://www.sphinx-doc.org/en/master/.

[168] [Online]. Available: https://mlbox.readthedocs.io/en/latest/.

[169] [Online]. Available: https://evalml.alteryx.com/en/stable/.

[170] F. T. K. a. Z. Z. Liu, "Isolation forest," in *8th ieee international conference on data mining*, Pisa, 2008.

[171] [Online]. Available: https://www.featuretools.com/.

[172] B. C. G. F. M. H. F. L. M. M. R. V. R. J. a. V. J. Bischl, "OpenML benchmarking suites and the OpenML100," 2017.

[173] B. C. G. F. M. G. P. H. F. L. M. M. R. v. R. J. a. V. J. Bischl, "OpenML Benchmarking Suites," 2019.

[174] [Online]. Available: https://github.com/shap/shap.

[175] [Online]. Available: https://aws.amazon.com/sagemaker.

[176] [Online]. Available: https://optuna.org/.

[177] [Online]. Available: https://github.com/ray-project/tune-sklearn.

[178] [Online]. Available: https://github.com/scikit-optimize/scikit-optimize.

[179] [Online]. Available: https://figshare.com/articles/dataset/PPG-BP_Database_zip/5459299.

[180] [Online]. Available: https://www.physionet.org/content/wearable-exam-stress/1.0.0/.

[181] [Online]. Available: https://www.physionet.org/content/simultaneous-measurements/1.0.2/.

[182] [Online]. Available: https://ubicomp.eti.uni-siegen.de/home/datasets/icmi18/.

[183] [Online]. Available: https://www.physionet.org/content/mmash/1.0.0/.

[184] [Online]. Available: https://www.physionet.org/content/eegmat/1.0.0/.

[185] [Online]. Available: https://www.physionet.org/content/ecg-spider-clip/1.0.0/.

[186] [Online]. Available: https://osf.io/2ax9d/.

[187] [Online]. Available: https://data.wprdc.org/dataset/municipal-building-energy-usage.

[188] [Online]. Available: https://springernature.figshare.com/collections/I-BLEND_a_campus_scale_commercial_and_residential_buildings_electrical_energy_dataset/3893581/1.

[189] [Online]. Available: https://stanfordnlp.github.io/contract-nli/.

[190] [Online]. Available: https://huggingface.co/datasets/pile-of-law/pile-of-law.

[191] [Online]. Available: https://github.com/lauramanor/legal_summarization.

[192] K. Maharana, S. Mondal and B. Nemade, "A review: Data pre-processing and data augmentation techniques," *Global Transitions Proceedings,* pp. 91-99, 2022.

[193] S. Mishra, P. K. Mallick, L. Jena and G.-S. Chae, "Optimization of skewed data using sampling-based preprocessing approach," *Frontiers in Public Health,* p. 555802, 2020.

[194] V. Werner de Vargas, J. A. Schneider Aranda, R. dos Santos Costa, P. R. da Silva Pereira and J. L. Victória Barbosa, "Imbalanced data preprocessing techniques for machine learning: a systematic mapping study," *Knowledge and Information Systems,* pp. 31-57, 2023.

[195] S. Cofre-Martel, E. Lopez Droguett and M. Modarres, "Big Machinery Data Preprocessing Methodology for Data-Driven Models in Prognostics and Health Management," *Sensors,* 2021.

[196] J. Wnek, "Constructive Induction," in *Encyclopedia of the Sciences of Learning*, Boston, MA, Springer US, 2012, pp. 781-783.

[197] S. Dhakal, S. Azam, K. M. Hasib, A. Karim, M. Jonkman and A. S. M. Farhan Al Haque, "Dementia Prediction Using Machine Learning," *Procedia Computer Science,* vol. 219, pp. 1297-1308, 2023.

[198] S. Y. a. M. F. Emily Black, "Fliptest: fairness testing via optimal transport," 2020.

[199] S. D. A. K. a. A. S. Diptarka Chakraborty, "Fair rank aggregation. Advances in Neural Information Processing Systems," p. 35:23965–23978, 2022.

[200] J. K. a. M. A. H. Marc P Hauer, ". Legal perspective on possible fairness measures–a legal discussion using the example of hiring decisions," in *Computer Law & Security Review*, 2021, p. 42:105583.

[201] V. A. R. G. D. H. J.-P. H. M. H. A. J. a. H. L. Florian Tramer, " Fairtest: Discovering unwarranted associations in data-driven applications," in *IEEE European Symposium on Security and Privacy (EuroS&P)*, 2017.

[202] T. D. B. Maarten Buyl, "Optimal transport of classifiers to fairness," 2022.

[203] K. S. a. G. K. Evaggelia Pitoura, "Fairness in rankings and recommenders: Models, methods and research directions," in *IEEE 37th International Conference on Data Engineering (ICDE)*, 2021.

[204] G. G. G. P. K. S. Dimitris Sacharidis, "Auditing for Spatial Fairness," in *EDBT*, 2023.

[205] K. T. G. G. D. S. E. P. N. T. D. R. D. F. I. Z. E. Loukas Kavouras, " Fairness Aware Counterfactuals for Subgroups," in *NeurIPS*, 2023.

[206] F. Amara, K. Agbossou, A. Cardenas, Y. Dubé, S. Kelouwani and others, "Comparison and simulation of building thermal models for effective energy management," *Smart Grid and renewable energy,* vol. 6, p. 95, 2015.

[207] A. Boodi, K. Beddiar, Y. Amirat and M. Benbouzid, "Building thermal-network models: A comparative analysis, recommendations, and perspectives," *Energies,* vol. 15, no. 4, p. 1328, 2022.

[208] E. Atam and L. Helsen, "Control-oriented thermal modeling of multizone buildings: Methods and issues: Intelligent control of a building system," *IEEE Control systems magazine,* vol. 36, no. 3, pp. 86-111, 2016.

[209] D. B. Crawley, L. K. Lawrie, F. C. Winkelmann, W. F. Buhl, Y. J. Huang, C. O. Pedersen, R. K. Strand, R. J. Liesen, D. E. Fisher and M. J. Witte, "EnergyPlus: creating a new-generation building energy simulation program," *Energy and buildings,* vol. 33, no. 4, pp. 319-331, 2001.

[210] M. Magni, F. Ochs, S. de Vries, A. Maccarini and F. Sigg, "Detailed cross comparison of building energy simulation tools results using a reference office building as a case study," *Energy and Buildings,* vol. 250, p. 111260, 2021.

[211] R. Kramer, J. Van Schijndel and H. Schellen, "Simplified thermal and hygric building models: A literature review," *Frontiers of architectural research,* vol. 1, no. 4, pp. 318-325, 2012.

[212] M. Lauster, J. Teichmann, M. Fuchs, R. Streblow and D. Mueller, "Low order thermal network models for dynamic simulations of buildings on city district scale," *Building and Environment,* vol. 73, pp. 223-231, 2014.

[213] G. Lilis, G. Giannakis, D. Rovas and E. Kosmatopoulos, "SRC and its applications to building thermal control," in *CLIMA 2013 - 11th REHVA World Congress and the 8th International Conference on Indoor Air Quality, Ventilation and Energy Conservation in Buildings*, Prague, 2013.

[214] K. Yang and J. Stoyanovich, "Measuring fairness in ranked outputs," in *29th international conference on scientific and statistical database management*, 2017.

[215] S. Caton and C. Haas, Fairness in Machine Learning: A Survey, New York, NY, United States: Association for Computing Machinery, 2024.

[216] "eQUEST: the QUick Energy Simulation Tool," 11 June 2024. [Online]. Available: https://www.doe2.com/equest/.

[217] B. Schölkopf, "Causality for Machine Learning," in *Probabilistic and Causal Inference*, ACM, 2022, p. 765–804.

[218] K. E. Willcox, O. Ghattas and P. Heimbach, "The imperative of physics-based modeling and inverse theory in computational science," *Nature Computational Science,* vol. 1, no. 3, pp. 166-168, 2021.

[219] A. Karpatne, G. Atluri, J. H. Faghmous, M. Steinbach, A. Banerjee, A. Ganguly, S. Shekhar, N. Samatova and V. Kumar, "Theory-guided data science: A new paradigm for scientific discovery from data," *IEEE Transactions on knowledge and data engineering,* vol. 29, no. 10, pp. 2318-2331, 2017.

[220] N. Sharma and Y. Liu, "A hybrid science-guided machine learning approach for modeling chemical processes: A review," *AIChE Journal,* vol. 68, no. 5, p. e17609, 2022.

[221] L. Von Rueden, S. Mayer, K. Beckh, B. Georgiev, S. Giesselbach, R. Heese, B. Kirsch, J. Pfrommer, A. Pick, R. Ramamurthy and others, "Informed machine learning-a taxonomy and survey of integrating prior knowledge into learning systems," *IEEE Transactions on Knowledge and Data Engineering,* vol. 35, no. 1, pp. 614-633, 2021.

[222] H. T. Su, T. J. McAvoy and P. Werbos, "Long-term predictions of chemical processes using recurrent neural networks: A parallel training approach," *Industrial & engineering chemistry research,* vol. 31, no. 5, pp. 1338-1352, 1992.

[223] Y. Tian, J. Zhang and J. Morris, "Modeling and optimal control of a batch polymerization reactor using a hybrid stacked recurrent neural network model," *Industrial & engineering chemistry research,* vol. 40, no. 21, pp. 4525-4535, 2001.

[224] J.-S. Chang, S.-C. Lu and Y.-L. Chiu, "Dynamic modeling of batch polymerization reactors via the hybrid neural-network rate-function approach," *Chemical Engineering Journal,* vol. 130, no. 1, pp. 19-28, 2007.

[225] W. K. Chan, B. Fischer, D. Varvarezos, A. Rao and H. Zhao, *Asset Optimization Using Integrated Modeling, Optimization, and Artificial Intelligence,* Google Patents, 2020.

[226] Z. Hao, S. Liu, Y. Zhang, C. Ying, Y. Feng, H. Su and J. Zhu, "Physics-informed machine learning: A survey on problems, methods and applications," *arXiv preprint arXiv:2211.08064,* 2022.

[227] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang and L. Yang, "Physics-informed machine learning," *Nature Reviews Physics,* vol. 3, no. 6, pp. 422-440, 2021.

[228] A.-p. Nguyen and M. R. Martínez, "MonoNet: Towards Interpretable Models by Learning Monotonic Features," *arXiv,* 2019.

[229] L. Breiman, "Random forests," *Machine learning,* vol. 45, pp. 5-32, 2001.

[230] . J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics,* pp. 1189-1232, 2001.

[231] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems,* vol. 30, 2017.

[232] A. Goldstein, A. Kapelner, J. Bleich and P. Pitkin, "Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation," *journal of Computational and Graphical Statistics,* vol. 24, no. 1, pp. 44-65, 2015.

[233] S. Wachter, B. Mittelstadt and C. Brent , "Counterfactual explanations without opening the black box: Automated decisions and the GDPR," *Harv. JL & Tech.,* vol. 31, p. 841, 2017.

[234] M. T. Ribeiro, S. Singh and C. Guestrin, "Anchors: High-precision model-agnostic explanations," in *Proceedings of the AAAI conference on artificial intelligence*, 2018.

[235] G. McLachlan and D. Peel, Finite Mixture Models, John Wiley & Sons, Inc., 2000.

[236] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE,* vol. 77, no. 2, pp. 257-286, 1989.

[237] J. Pearl, "Bayesian networks: A model cf self-activated memory for evidential reasoning," in *Proceedings of the 7th conference of the Cognitive Science Society, University of California*, Irvine, CA, USA, 1985.

[238] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, “Generative adversarial nets,” *Advances in neural information processing systems,* vol. 27, 2014.

[239] D. P. Kingma and M. Welling, “Auto-Encoding Variational Bayes,” *arXiv preprint arXiv:1312.6114,* 2013.

[240] D. Rezende and S. Mohamed, “Variational inference with normalizing flows,” in *International conference on machine learning*, 2015.

[241] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan and S. Ganguli, “Deep unsupervised learning using nonequilibrium thermodynamics,” in *International conference on machine learning*, 2015.

[242] J. Ho , A. Jain and P. Abbeel, “Denoising Diffusion Probabilistic Models,” *arXiv preprint arXiv:2006.11239.,* 2020.

[243] V. Borisov , K. Seßler , T. Leemann , M. Pawelczyk and G. Kasneci, “Language Models are Realistic Tabular Data Generators,” *arXiv preprint arXiv:2210.06280,* 2023.

[244] “Living guidelines on the responsible use of generative AI in research,” 20 March 2024. [Online]. Available: https://research-and-innovation.ec.europa.eu/document/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en.

[245] “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,” 2021 April 21. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206.

[246] “Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on liability for defective products,” 28 September 2022. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52022PC0495.

[247] “Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive),” 28 September 2022. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52022PC0496.

[248] “Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act),” 30 May 2022. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2022/868/oj.

[249] “Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act),” 13 December 2023. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2023/2854/oj.

[250] S. Klein, “University of Wisconsin-Madison solar energy laboratory,” *TRNSYS: A transient simulation program. Eng. Experiment Station,* 1975.

[251] K. K. Vu, C. D'Ambrosio, Y. Hamadi and L. Liberti, “Surrogate-based methods for black-box optimization,” *International Transactions on Operational Research,* pp. 393-424, 2016.

[252] “User Story Mapping - tarefas com foco no usuário,” 20 August 2019. [Online]. Available: homemmaquina.com.br/user-story-mapping/.

[253] "RapidMiner," Altair, [Online]. Available: https://altair.com/altair-rapidminer. [Accessed 5 4 2024].

[254] "Power BI," [Online]. Available: https://www.microsoft.com/en-us/power-platform/products/power-bi/.

[255] "Scikit-learn," [Online]. Available: https://scikit-learn.org/stable/.

[256] Z. Fan, H. Fang, Z. Zhou, J. Pei, M. P. Friedlander, C. Liu and Y. Zhang, "Improving Fairness for Data Valuation in Horizontal Federated Learning," in *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, 2022.

[257] A. Ghorbani and J. Zou, "Data shapley: Equitable valuation of data for machine learning," in *International conference on machine learning*, 2019.

[258] J. T. Wang and R. Jia, "A Note on "Towards Efficient Data Valuation Based on the Shapley Value"," *arXiv:2302.11431,* 2023.

[259] J. T. Wang and R. Jia, "Data Banzhaf: A Robust Data Valuation Framework for Machine Learning," in *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics*, 2023.

[260] N. Dong, X. Zhou, t. Xiao and S. Sun, "Industry Electricity Demand Forecast by Integrating Multiple Correlation Analysis Methods," in *2021 IEEE 5th Conference on Energy Internet and Energy Systems Integration (EI2)*, 2021.

[261] "Apache Spark - A Unified engine for large data analytics," Apache, [Online]. Available: https://spark.apache.org/docs/3.5.3. [Accessed 11 November 2024].

[262] "What is Amazon Kinesis Data Streams?," 11 November 2024. [Online]. Available: https://docs.aws.amazon.com/streams/latest/dev/introduction.html.

[263] A. Figueira and B. Vaz, "Survey on synthetic data generation, evaluation methods and GANs," in *Mathematics 10 (15)*, 2022, p. 2733.

[264] "AI Fairness 360 - IBM," [Online]. Available: https://aif360.res.ibm.com/.

List of Acronyms/Abbreviations

Acronym/ Abbreviation	Description
DoA	Description of Action
MVP	Minimum Viable Product
WP	Working Package