H2020 - INDUSTRIAL LEADERSHIP - Information and Communication Technologies (ICT)

ICT-38-2020: Artificial intelligence for manufacturing

Grant Agreement: 957362

Explainable Manufacturing Artificial Intelligence

XMANAI

MAKING AI UNDERSTANDABLE

WP1: Explainable AI Foundations Elaboration and XMANAI Concept Fusion

D1.4: Final XMANAI MVP

Deliverable Leader: Suite5

Due Date: 31/01/2023

Dissemination Level: Public

Version: F1.0

Short Abstract

This deliverable aims at bringing together the XMANAI concept by: (a) describing the different user journeys in Explainable AI for business users, data scientists and data engineers, (b) eliciting the backlog of technical requirements and aligning them with the business requirements and the user journeys, (c) consolidating the Minimum Viable Product (MVP) that summarizes the expected features on which XMANAI shall focus (by the end of the project) for maximizing the expected added value to manufacturers while ensuring innovation from a scientific and technical perspective; and (d) defining the technical aspects regarding the concept of "X-By-Design" (Explainable by Design). It takes into account the feedback and experiences gained through the development and integration activities of the alpha release of the XMANAI Platform, as well as the initial implementation activities of the XMANAI Demonstrators, in order to finalise the specifications of the XMANAI solution.

Further Information: www.ai4manufacturing.eu

Document Log

Contributors	**Suite5**, TXT, Fraunhofer, TYRIS, POLIMI, AIDEAS, INNOVALIA, ATHENA, KBIZ, WHR, FORD, UBITECH, DBL, UNIMETRIK, CNH
Internal Reviewer 1	TYRIS
Internal Reviewer 2	TXT
Type	Report
Delivery Date	12/05/2023 (M31)

History

Versions	Description
D0.1	Initial Table of Contents
D0.2	Initial draft of Section 2, 4 consolidating the feedback of the implementation activities in WP2-WP5
D0.3	Initial draft of Section 3 containing inputs by technical partners
D0.4	Initial draft of Section 5 containing inputs by demo partners
D0.5	Full Deliverable Draft sent for internal review
R0.1	Revision of internal reviewer 1 (TYRIS)
R0.2	Revision of internal reviewer 2 (TXT) based on R0.1
D0.8	Updated version addressing comments received during the internal review
F1.0	Final version submitted to the EC

Executive Summary

The XMANAI Deliverable D1.4 “Final XMANAI MVP” documents the results achieved by tasks T1.3 “Platform Requirements Elicitation, Data Acquisition and AI Scenarios” and T1.4 “XMANAI Concept Elaboration, MVP Definition and Validation”, as well as the performed work in this context. The purpose of this deliverable is to consolidate the basic ingredients of the XMANAI concept in the form of user journeys, technical requirements and features of the Minimum Viable Product (MVP) that has driven the implementation steps of the project.

Following the patterns of the agile development methodology, the XMANAI high-level usage scenarios, noted as User Journeys, have been elaborated considering three different, but equally important, roles: the Business User, the Data Scientist and the Data Engineer. These scenarios define the TO-BE situation for each role in order to highlight the current challenges and the expected changes that shall be introduced through XMANAI. For each TO-BE User Journey, different phases (in total 8, across the different stakeholders) have been described along with each stakeholder’s expectations, benefits and challenges.

In terms of capturing the actual users’ needs, such user journeys are complemented by the collection of 114 technical requirements that have been brainstormed and “groomed” by all partners in the form of user stories and are accompanied by 19 non-functional requirements. The mapping between the technical requirements and: (a) the different steps envisaged across the phases of the users journeys, and (b) the 46 updated business requirements directly gathered by the project’s demonstrators (as elaborated in the XMANAI Deliverable D6.3), has been explicitly provided.

Finally, the process and the final outcomes of the XMANAI Minimum Viable Product (MVP) have been defined. Taking into consideration the internal and external assessment provided in the draft and intermediate MVP as well as the latest project’s advancements in the XMANAI Platform implementation and the XMANAI Demonstrators’ operation, the final MVP release has eventually led to 36 “Will-Have” features (or epics), and 9 “Won't-Have” features (that essentially inherit their prioritization to their associated technical requirements).

The results of this deliverable have been constantly aligned with the XMANAI architecture design in “T5.1- Platform Architecture, Bundles Communication Design and APIs Definition”, as well as with the design, development and integration activities in WP2 “Industrial Asset Management and Secure Asset Sharing Bundles”, WP3 “Core Artificial Intelligence Bundles for Algorithm Lifecycle Management”, WP4 “Novel Artificial Intelligence Algorithms for Industrial Data Insights Generation” and WP5 “XMANAI Platform Continuous Integration”. Feedback received during the implementation of the alpha version of the XMANAI integrated platform as well as the preparatory and early implementation activities for the alpha phase of the XMANAI demonstrators has been leveraged in the final MVP consolidation.

Table of Contents

List of Figures

List of Tables

1 Introduction

The main aim of this section is to provide a brief overview of the deliverable.

1.1 XMANAI Project Overview

Despite the indisputable benefits that Artificial Intelligence (AI) can bring in society and in any industrial activity, humans typically have little insight about AI itself and even less concerning the knowledge on how AI systems make any decisions or predictions due to the so-called “black-box effect”. Many of the machine learning/deep learning algorithms are opaque and not possible to be examined after their execution to understand how and why a decision has been made. In this context, to increase trust in AI systems, XMANAI aims at rendering humans (especially business experts from the manufacturing domain) capable of fully understanding how decisions have been reached and what has influenced them.

Building on the latest AI advancements and technological breakthroughs, XMANAI shall focus its research activities on Explainable AI (XAI) in order to make the AI models, step-by-step understandable and actionable at multiple layers (data-model-results). The project will deliver "glass box" AI models that are explainable to a "human-in-the-loop", without greatly sacrificing AI performance. With appropriate methods and techniques to overcome data scientists’ pains such as lifecycle management, security and trusted sharing of complex AI assets (including data and AI models), XMANAI provides the tools to navigate the AI’s “transparency paradox” and therefore:

(a) accelerates business adoption addressing the problematic that “if manufacturers do not understand why/how a decision/prediction is reached, they will not adopt or enforce it”, and
(b) fosters improved human/machine intelligence collaboration in manufacturing decision making, while ensuring regulatory compliance.

XMANAI aims to design, develop and deploy a **novel Explainable AI Platform** powered by explainable AI models that inspire trust, augment human cognition and solve concrete manufacturing problems with value-based explanations. Adopting the mentality that “AI systems should think like humans, act like humans, think rationally, and act rationally”, a catalogue of **hybrid and graph AI models** is built, fine-tuned and validated in XMANAI at 2 levels: (i) baseline AI models that will be reusable to address any manufacturing problem, and (ii) trained AI models that have been fine-tuned for the different problems that the XMANAI demonstrators’ target. A bundle of **innovative manufacturing applications and services** are also built on top of the XMANAI Explainable AI Platform, leveraging the XMANAI catalogue of baseline and trained AI models.

XMANAI will validate its AI platform, its catalogue of hybrid and graph AI models and its manufacturing apps in **4 realistic, exemplary manufacturing demonstrators** with high impact in: (a) optimizing performance and manufacturing products’ and processes’ quality, (b) accurately forecasting product demand, (c) production optimization and predictive maintenance, and (d) enabling agile planning processes. Through a scalable approach towards Explainable and Trustful AI as dictated and supported in XMANAI, manufacturers will be able to develop a robust AI capability that is less artificial and more intelligent at human and corporate levels in a win-win manner.

1.2 Deliverable Purpose and Scope

This deliverable documents the final results of tasks T1.3 “Platform Requirements Elicitation, Data Acquisition and AI Scenarios” and T1.4 “XMANAI Concept Elaboration, MVP Definition and Validation” for the final iteration of WP1 “Explainable AI Foundations Elaboration and XMANAI Concept Fusion”.

Its main purpose is to elaborate in detail on the XMANAI concept as a novel approach for using explainable AI in manufacturing operations, as well as to provide an up-to-date version of the XMANAI Minimum Viable Product (MVP) that has been used as a guide for the research and development activities of the project since its very beginning. To achieve these goals, the document has provided

the key ingredients in line with its predecessors, i.e. the XMANAI Deliverables D1.2 and D1.3, adapted to incorporate the latest research findings and follow the project's advancements.:

a) The User Journeys for all involved roles, presenting the TO-BE processes along with the expected benefits and challenges,
b) The updated Technical Requirements regarding the XMANAI Platform, in the form of User Stories, and
c) The list of the most valuable and feasible features that form the final XMANAI MVP.

1.3 Impact and Target Audiences

The results presented in this document target mainly the technical users that develop the XMANAI Platform, as well as the researchers who support the solution. The business users are also influenced by the content, since in essence it describes the XMANAI solution that is under development.

The resulted MVP is meant to act as the driving force behind the technical discussions and become the core technical blueprint on top of which the design of the architecture and the implementation of the platform has been based.

1.4 Deliverable Methodology

The information reported in the series of deliverables regarding the requirements and the MVP has been produced by the consortium members following the methodology described below:

I. **User Journeys Revision**: In D1.2, User Journeys had been derived from repeated brainstorming sessions with the stakeholders using Miro boards[1]. The workflows that had been put together for each of the high-level user roles (Business User, Data Engineer, Data Scientist) have been refined in this deliverable, taking into consideration the XMANAI Platform-oriented User Journeys described in D5.2.
II. **Technical Requirements Refinements**: In D1.2, the technical requirements had been derived, on the same fashion through brainstorming using the Miro boards, with repeated iterations with the technical partners of the consortium. The consolidated list of user stories has been maintained and kept up-to-date in D1.3 and in this deliverable in alignment with the project's development activities (in WP2, WP3, WP4 and WP5).
III. **MVP Finalization**: In the draft and intermediate MVP, the wide range of technical requirements had been grouped and consolidated into different MVP features, that had been assessed for: (a) their added value (by demonstrator partners and external stakeholders/members of the XMANAI Industrial Advisory Board), and (b) their feasibility (by technical partners) in order to produce a prioritized list of features and functionalities. In the final MVP, the list of MVP features that will be implemented (Will-Have) vs what will not be implemented (Won't Have) has been defined with clarity.

It needs to be noted that the data requirements that were described in the XMANAI Deliverables D1.2 and D1.3 through the available data sources, their features and limitations, as well as the means of acquisition for the respective demonstration scenario, have remained essentially the same, so it was considered as prudent not to repeat the same information in this deliverable. In fact, with the ongoing demonstration activities in WP6, the actual datasets have been already uploaded in the XMANAI Platform and their profile (structure and metadata) is available online.

1.5 Dependencies in XMANAI and Supporting Documents

As already stated, the XMANAI Deliverable D1.4 reports the results of Tasks T1.3 "Platform Requirements Elicitation, Data Acquisition and AI Scenarios" and T1.4 "XMANAI Concept Elaboration, MVP Definition and Validation" of WP1 "Explainable AI Foundations Elaboration and XMANAI Concept Fusion". Both of these tasks make use of the outcome of the tasks T1.1 "Explainable AI and Graph

[1] https://miro.com

Machine Learning Analytics State-of-Play" and T1.2 "Human Aspects in Decision Making and AI" (reported in D1.1). Task 1.3, in particular, also compares the technical requirements gathered from technical partners, with the latest business requirements collected in the framework of T6.1 "Demonstrators Requirements Elicitation".

More specifically, this deliverable is expected to complement the architectural design task in "T5.1-Platform Architecture, Bundles Communication Design and APIs Definition", and the integration towards the different releases of the XMANAI Platform in WP5. It has been constantly aligned with the implementation of all "WP2 – Industrial Asset Management and Secure Asset Sharing Bundles", "WP3 - Core Artificial Intelligence Bundles for Algorithm Lifecycle Management" and "WP4 - Novel Artificial Intelligence Algorithms for Industrial Data Insights Generation" activities. It needs to be noted that the consolidated MVP has also taken into consideration the feedback from the XMANAI demonstrators' operation in their alpha release in WP6.

1.6 Document Structure

The contents of this document are structured in sections as follows:

- Section 2 presents the XMANAI workflows through the concept of User Journeys and is divided in phases according to the user Role (Business User, Data Engineer, Data Scientist), with each phase having its own specific steps and features.
- Section 3 focuses on the final technical requirements as expressed by business experts and data science professionals of the consortium partners in the form of User Stories.
- Section 4 presents the consolidation of the final XMANAI Minimum Viable Product (MVP).
- Section 5 introduces the levels of explainability and the technical aspects that pertain to the concept of "Explainable-By-Design" in XMANAI.
- Section 6 concludes this deliverable.

2 User Journeys

This section is dedicated to the different user journeys that were initially introduced in the XMANAI Deliverables D1.2-D1.3 and have been updated to reflect the progress of the development activities of XMANAI and the feedback received by the demonstration activities with the end users and the external stakeholders in the industrial domain.

2.1 Stakeholders

The different stakeholders that are expected to utilize the XMANAI project results focusing on Explainable AI for the manufacturing industry, are classified into 3 high-level categories:

- **Business Users**: Domain experts within a manufacturer that specialize into the operations of different departments (e.g. production, marketing, sales, IT, etc.) and who need to understand the results of an analysis in a tangible - for them - manner. Business Users are also responsible for assessing the validity of the results and explanations from a business perspective. In XMANAI, the targeted business users typically vary per demonstrator, as described in detail in Deliverable D6.1, D6.3 and D6.4.
- **Data Scientists**: Data experts that use scientific methods, processes, and algorithms to explore and extract knowledge from data, understand the problem at hand, create AI models and derive actionable insights in different application domains. Data Scientists are also responsible for ensuring - from a scientific perspective - the trustworthiness and robustness of the results, both the predictions per se and their associated explanations.
- **Data Engineers**: Software engineering experts with a strong data background that are responsible for building the necessary underlying infrastructure to collect and prepare data, and to deploy AI models and solutions to analyze data in a scalable manner. Data engineers also ensure the integrity of the data and all their associated security aspects.

2.2 Business User Journey in XMANAI

The initial Business User journey was created collaboratively by the XMANAI partners through an exercise using Miro boards (as described in the XMANAI Deliverable D1.2) and consists of three phases: Phase 1, related to how data can be provided and communicated in an efficient way, from the business user perspective, towards XAI Preparation. Phase 2, that deals with the iterative understanding and evaluation of XAI models and analytics results for AI experimentation, and Phase 3, which refers to the insights gained by the XAI results and how these can increase business efficiency while XAI is in “production”.

These 3 phases have been modified to accommodate the changes and refinements produced during the final WP1 iteration, taking into consideration the updated User Journeys described in the XMANAI Deliverable D5.2, from a platform perspective.

Phase 1: XAI Preparation - Provide and Understand Data

The XAI journey for the Business User (with a strong understanding of the data within the manufacturing domain) commences with the provision of the manufacturing data that have been already prepared offline for the task required. The user utilizes the XMANAI platform and configures a secure data harvesting process, that can be realized either through an API or by uploading files in tabular format (e.g. csv files). The configuration includes the mapping of the data to the XMANAI data model, and the definition of the appropriate access policies that will be enforced every time a stakeholder may attempt to access the specific data. For semantic, exchange (sharing) and search purposes, the user is expected to provide also the metadata of the data asset which consist of a description, various data characteristics and licensing details among other things.
The execution of the data harvesting service may be performed in the cloud (in dedicated spaces per organization) or in an on-premise environment (private cloud installation or on-premise servers) depending on the specific manufacturer’s preference and can be configured to occur immediately or

according to a schedule (for API harvesting). On a successful execution, the data are stored in the XMANAI platform and from this point any changes performed on this asset are tracked. If the execution fails, the information related to the error is collected in order to help the business user to understand what went wrong and correct it. It needs to be noted that if any dataset contains sensitive information, it should be anonymized in the on-premise environment first.

During the same phase, the business user may as well access the asset marketplace of the XMANAI platform in order to find other data assets of interest that will help his/her cause. If the user discovers such an asset, he/she can request access to it through a secure sharing mechanism, which is formalized via a valid data asset agreement. Once an asset is acquired, it can be immediately used in the XMANAI platform for a time period designated in the contract terms.

This preparation phase, at its current status, aims to adhere to the main expectations of the business users regarding the XMANAI platform, which can be summarized into:

- Data protection in terms of a secure data management environment where data are easily ingested, while privacy is protected.
- Data interoperability, considering both its semantic and syntactic perspectives, where data are expected to be able to be exchanged and integrated in an efficient manner.
- Data sharing to ensure that all involved stakeholders (e.g. demonstrator and technical partners in the context of the XMANAI project) can have access to the data and all its accompanying information in an IPR-respectful manner and considering the access policies that the business users have defined.

Moreover, the main challenges of this phase, as presented in D2.1-D2.3, are addressed, especially those that concern: (a) effective handling of connection, synchronization and interoperability issues across the various data sources; (b) dealing with protection and privacy regulations; (c) providing appropriate semantics in datasets in the most efficient manner.

Table 2-1: Business User Journey Phase 1 - XAI Preparation

Phase 1: AI Preparation - Provide and Understand Data	
Actions	Ingest data in different ways (e.g. through APIs or as csv files)
	Map data to a common data model
	Annotate data with column definitions or data types through the use of the data model
	Add metadata on each data source
	Anonymise data to protect privacy, if required, using on premise XMANAI anonymiser tool
	Define access policies and privacy constraints for each dataset
	Keep history log for each dataset's life cycle
Challenges Addressed	Handle access to confidential data by storing data on premise (i.e. on private cloud infrastructures, on private servers).
	Ensure an effective data collection, connection and synchronisation of various, possibly heterogeneous, data sources by using a common data model
	Consider data protection and privacy challenges (GDPR) using specific access polices as metadata and allowing anonymisation of sensitive information
	Deal with low quality or missing metadata by enabling the user to improve the quality of metadata manually (i.e. edit metadata)

Phase 2: XAI Experimentation - Contribute to better understanding and evaluating the XAI models / pipelines / results

The second phase of the Business User journey refers to the process of providing business/domain knowledge in order to support data scientists and the creation of appropriate analytics models, the evaluation of ML/DL models and their results.
In order to contribute the most, the business user needs to have a solid understanding of the business problem at hand. They can do so, by accessing the available XAI results during experimentation. At the same place, they can view related visualisations or summary statistics, so as to have a clear picture before they provide their feedback through annotations and comments in the related XAI pipelines, as provided by the XMANAI platform. Moreover, the performance of the pipelines and their XAI models can be tracked with the selected evaluation metrics.
The insights gained in the XMANAI platform fulfil the expectations of business users regarding this phase. In addition, they assist them to understand the reasoning behind XAI results (including predictions, explanations, performance) and reliably tell whether a presented XAI model or pipeline is successful or requires further experimentation. If satisfied, they can decide to integrate the given pipeline in production and their actual operations.
The view of reliable and comprehensive results during the training and testing phase of XAI models increases the trust of business users towards AI, which is a core challenge for XMANAI.

Table 2-2: Business User Journey Phase 2 - XAI Experimentation

Phase 2: XAI Experimentation - Contribute to better understanding and evaluating the data / XAI models / results	
Actions	View the selected performance metrics for each XAI model/pipeline and provide feedback
	Review results/performance/explanations of various executions of XAI models/pipelines
	Compare results/performance/explanations of various XAI models/previous runs
	Easily identify unfit/flawed XAI models
	Provide appropriate visualisations to easily identify the results
	Provide domain knowledge (descriptions, tags, relations, business objectives) for better explaining results
	Ask exploratory questions over the training phase or the testing results
	Give targeted feedback on results and explanations (based on domain knowledge)
Challenges Addressed	Create the conditions for efficient communication among Business Users, IT teams and Data Scientists by providing the appropriate communication tools and by allowing all roles to give/receive feedback on the XAI models/pipelines
	Make business users understand the XAI results and what XAI models are doing by providing simple visualisations, statistics and explained performance metrics
	Have a simple and easy interaction with the platform

Phase 3: XAI Insights - Understand XAI results

Gaining appropriate insights into XAI results while an XAI system is in production is the third and last phase of this journey. A business user should be provided with proper tools that enhance the explainability of the results, monitor risk assessment and relate the outcomes with real business KPIs.
During this phase, the business user can study the security/risk assessment of the model/pipeline involved and receive useful insights and visualizations concerning the explainability of the utilized

models. In this manner, explainability becomes the main tool for data and AI outcome quality assessment. Extracting tangible and actionable knowledge from XAI results builds a circle of trust between AI and business users and increases the adoption of AI in their everyday work without excluding the human factor.

Table 2-3: Business User Journey Phase 3 - XAI Insights

Phase 3: XAI Insights - Understand XAI results	
Actions	Receive explainable results in an appropriate way (e.g. through visualisations, as text)
	Run what-if scenarios with different key parameters in XAI models
	Monitor risk/security assessment per XAI model
	Define actual business KPIs to which XAI results can be related
Challenges Addressed	Relate XAI results to business processes to help business users understand better the outcome through visualisations and explainability
	Make clear that this promising AI technology will not replace existing methods/ human experience, but it will work in a complementary fashion and help the business users perform more efficiently

2.3 Data Scientist Journey in XMANAI

The initial Data Scientist User journey was created collaboratively by the XMANAI partners through an exercise using Miro boards (in XMANAI Deliverable D1.2) and consists of five phases: Phase 1, related to understanding the available data and the problem that needs to be handled as an XAI preparatory phase. Phase 2, that covers the data preparation process including the proper manipulation of the problematic data cases. Phase 3, where the design of the pipeline takes place and the related XAI models are trained. In this process, data scientists should be able to collaborate and share their ideas and configurations with other users. Phase 4 involves all the actions that are required so that the results and the models themselves are explained, in order to allow different users to comprehend how the outputs occurred. Finally, Phase 5 refers to the evaluation methods for assessing the performance of the models, so that the changes that will improve the performance can be recognized and made.

These 5 phases have been modified to accommodate the changes and refinements produced during the final WP1 iteration, taking into consideration the updated User Journeys described in the XMANAI Deliverable D5.2, from a platform perspective.

Phase 1: XAI Preparation - Understand the data and the problem at hand

During phase 1, data scientists get access to the data available under their organization (either owned by their organization or shared with their organization) and get familiar with them in order to better understand all the aspects of the problem at hand. They can also search for additional datasets based on various metadata features, view their details and request access. Similar to the business user, the data scientist can also initiate and conclude a sharing agreement process with a data owner. Once the data are accessible, the data scientist may perform data exploration through visualizations and statistical reports of the data or by the data profile as described by the asset metadata (e.g. features, datatypes, number of samples, etc). These semantics also provide the means for knowledge transfer from the business users towards the data scientists. In this way, data scientists are able to quickly access necessary information to comprehend the data and the problem, which was one of the main challenges as recorded in D1.2.

After completing phase 1, data scientists have a clear view of the data preparation steps required to get the data ready for model experimentations (i.e. training and testing) and proceed to phase 2.

Table 2-4: Data Scientist Journey Phase 1 – XAI Preparation

Phase 1: XAI Preparation – Understand the data and the problem at hand	
Actions	Find the appropriate datasets for the specified task. Request for more data through the marketplace.
	Acquire domain knowledge about the data, terminology and processes of the problem at hand by studying the metadata
	Study the nature of data in terms of volume (number of samples, number of features), velocity (speed of generation and collection) and variety (data types, timeseries, tabular etc), type of features (continuous, categorical, mixed)
	Perform data exploration to get quality and availability information, visualizations, correlations, statistics
	Search for relevant additional datasets that are available based on a dataset's metadata
Challenges Addressed	Understand the problem at hand using various data exploration techniques
	Achieve efficient communication with the business user to get help with understanding the data through metadata information

Phase 2: XAI Experimentation - Prepare the data and handle problematic data cases

During phase 2, the data scientist performs a number of data manipulation techniques to get the data ready before training the models. More specifically, data scientists need to convert all data to common formats and match / join different datasets from different data sources. By performing this data harmonization step, data from different sources are made interoperable and comparable. The functionalities for data transformation, outlier detection and missing data handling operations are also available through the XMANAI data preparation functionalities. In addition, the data scientist can easily perform filtering, merging, normalization and scaling. Through visualizations, data scientists are able to validate the outcomes of each data preparation step. In cases where data quantity is not adequate or class imbalance is present, data scientists can choose to perform a data augmentation technique (e.g. add modified copies of existing data or new synthetic samples). Finally, the data preparation pipeline is stored, so that data scientists can easily re-apply the same data preparation steps to new incoming data.

Since data preparation is considered to be the most challenging and time-consuming phase in any data analytics project, a lot of effort needs to be placed in XMANAI to ensure that data from various sources can be manipulated, preprocessed and transformed in a consistent and efficient manner. Making sure that pipelines can be re-used also greatly reduces the overall time of the data preparation phase.

Table 2-5: Data Scientist Journey Phase 2 – AI Experimentation

Phase 2: AI Experimentation - Prepare the data and handle problematic data cases	
Actions	Perform data harmonization to convert all data to common formats (making data from different sources interoperable and comparable among them)
	Find feature correlations and preliminary feature relevance to the task in order to keep only the appropriate features for the models
	Perform data transformation operations (basic math, change data types, etc.), missing value handling operations (imputation), handling outliers operations (drop, assign a new value)
	Perform more advanced transformation operations like normalization (feature scaling), data integration (e.g. appending rows or appending columns), filtering
	Make use of data augmentation techniques

	Recognize and deal with class imbalance (oversampling, down-sampling, synthetic samples)
Challenges Addressed	Data preparation in XMANAI does not take too much of a data scientist time since all of the required tools are available for easy usage and quick visualisations for each step's outcome are provided
	Ensured the same data processing and transforming pipeline can be applied in the inference phase since pipelines can be stored as autonomous entity
	Build a dataset representative of the problem that captures the variability and the problem dynamics by using the data preparation available tools
	Combining data of different types and data sources by transforming data to the same format and then merge them

Phase 3: XAI Experimentation - Collaboratively Design XAI pipelines

Phase 3 involves all the steps required for the design, training and evaluation of XAI flows called pipelines, which usually consist of data preparation routines and XAI models (pre-trained or not).
During design time, the data scientists can work on different XAI models by selecting the most appropriate algorithms and performing experiments to decide on the best model and its optimal parameterisation to be employed. The experiments are kept and maintained in an experimentation tracking environment available to all users of the organisation, where results, parameters and performance metrics can be evaluated and compared. When the data scientist is satisfied with the evaluation and qualifies a model, or if there is already a pre-trained model available, he/she can proceed to design an XAI pipeline.
For this purpose, the XMANAI platform provides a user-friendly UI to design, manage and monitor XAI pipelines and an extensive catalogue of ready-to-use XAI models, including both baseline algorithms and pre-trained models. Data scientists are able to design, experiment, train and tune the performance of as many pipelines and models as needed. Each stage of the pipeline can be implemented/modified independently of other stages, so that the users are able to split the implementation workload and benefit from modular agile development. Since the feedback of other users is vital, the pipelines can be available to other XMANAI roles (scientists, engineers, business) of the same organization too. Before sharing the final pipeline, the data scientists can annotate and explain the overall pipeline results.
All these platform features assist immensely the work of a data scientist and address most of the challenges usually encountered in this domain. The efficient creation and update of XAI pipelines along with the ability to compare results and collaborate with other stakeholders at this phase, help data scientists to meet faster their own requirements and build pipelines that meet also the requirements of the business users.

Table 2-6: Data Scientist Journey Phase 3 – XAI Experimentation

Phase 3: XAI Experimentation - Collaboratively Design XAI pipelines	
Actions	Split development workload to independently implement/modify stages of the pipeline (thanks to modular agile development)
	Build and experiment with as many pipelines and models as needed
	Train, monitor and tune XAI pipelines performance
	Save the pipelines (trained or not) for later usage
	Share the pipelines with other data scientists to get feedback about any improvements on the chosen approach
	Share the pipelines with data engineers to get feedback on designing a production ready pipeline

	Share the pipelines and results with business users to get feedback on which outputs to pay more attention to
Challenges Addressed	Lack of reproducibility / modularity is addressed by storing pipelines as reusable elements.
	Choosing between complex models or simpler ones becomes easy when all models can be viewed together and be compared.
	Limited collaboration with other users (especially business users) is alleviated by sharing pipelines and experimental results with explanations and allow for feedback

Phase 4: XAI Experimentation - Explain XAI models / results

In this phase, data scientists need to use appropriate explainability techniques to explain the results obtained from the XAI models. The XMANAI platform provides the data scientists with popular explainability approaches able to highlight the reasons and get insights on how the results occurred (e.g., relevance of inputs). According to the nature of the problem, they can decide which explainability technique fits best the task and create simple visualizations to effectively communicate the results to the business users.

By explaining the results of the AI models, data scientists can get insights about the models' inner workings and the way their decisions are taken, inspect for biases in the results and redesign the pipeline if needed. They are also able to avoid obvious outputs as meaningful explanations of the XAI models. In addition, understanding the way the models function and produce the outputs increases the confidence for the accuracy and the robustness of the models. Business users are also benefited from the explanations and the visualizations produced, as the information they receive can help them take beneficial decisions for their business operations.

In this direction, the XMANAI platform has focused on matching the expectations of both the data scientist and the business user concerning explainability which require transparency, simplicity as well as deep understanding of the reasons that produced the output.

Table 2-7: Data Scientist Journey Phase 4 –X AI Experimentation

Phase 4: XAI Experimentation - Explain XAI models / results	
Actions	Determine the explainability method that fits better to the problem by checking the needs of the business users
	Create visualizations to better convey the model results and the decisions to the business users
	Inspect for bias in the results and avoid obvious outputs as explanations
Challenges Addressed	Difficulty in making explanations useful to the business user (e.g., in the form of rules)
	Selecting the explainability techniques that match the expectations of the business user per problem
	Breaking down the complexity of black box models
	Lack of available explainable AI libraries which are consistent and robust

Phase 5: XAI Experimentation - Evaluate XAI models / pipelines

During phase 5, data scientists monitor the models' performance using different metrics and validation methods. The XMANAI platform provides an experimentation environment with the option to choose different performance metrics according to the model in use (e.g., different metrics for regression versus classification models), so that the right metrics will be used for the validation process. In addition, data scientists can easily apply the chosen validation method and even

incorporate it in the XAI pipeline (e.g., cross validation approach, hold out, k-fold cross validation). Common visualization libraries for visualizing the performance metrics of an experiment are also available, as well as the ability to log experiments (model runs) & metrics and be able to compare them at will. It is critical in this phase to evaluate not only the accuracy of the results, but also the validity and understanding of the explanations for the target business users, by appropriately involving them.

The evaluation phase of XAI models gives data scientists the opportunity to debug and improve the performance of the trained models. They can choose the most appropriate, explainable, reliable and accurate model (best model parameters), by monitoring and visualizing the performance metrics of different configurations and architectures. Data scientists can also recognize the most relevant features through the evaluation procedure and use them for a potential retraining in the future.

Table 2-8: Data Scientist Journey Phase 5 – XAI Experimentation

Phase 5: Evaluate XAI models / pipelines	
Actions	Define the proper validation method according to the problem
	Choose the right performance metrics that will be used for the validation process
	Log the results of different experiments and compare the performance of the different models in a common experimentation environment
	Create visualizations to efficiently present the validation results in a graphical format
Challenges Addressed	Able to choose appropriate model evaluation techniques
	Able to find the right metrics to attach importance to, in agreement with the needs of the business users

2.4 Data Engineer Journey in XMANAI

The initial Data Engineer User journey was created collaboratively by the XMANAI partners through an exercise using Miro boards. It consists of three phases: Phase 1, related to the ingestion and handling of data in a collaborative way. Phase 2, that involves the design of XAI pipelines in collaboration with data scientists and Phase 3, that deals with the deployment of the XAI pipelines to production.

These 3 phases have been modified to accommodate the changes and refinements produced during the final WP1 iteration, taking into consideration the updated User Journeys described in the XMANAI Deliverable D5.2, from a platform perspective.

Phase 1: XAI Preparation - Collaborate on uploading / handling data

During phase 1, data engineers need to handle the data ingestion process in collaboration with the business users. More specifically, the data engineers can use the XMANAI platform to create a new data harvesting process and configure it based on the nature of the data i.e., source, format, velocity, volume, intended usage, or directly upload the relevant data files. In addition, during ingestion they can semantically map the data to the XMANAI data model in order to ensure the data are understandable to any stakeholder that will utilize them.

Data engineers are also given the ability to define data handling processes as needed, such as the definition of access policies and other metadata or the setup of anonymisation rules. Finally, they can schedule an (repeatable or not) execution time, associated to the data asset, for API harvesting or perform the execution immediately (usually for files). They can also define whether the relevant service is triggered for execution in the cloud (in dedicated spaces/VMs per organization) or in an on-premise environment (e.g. private cloud installation). If the execution is successful, the data are stored

in the XMANAI platform, where data lineage and versioning tools are applied so that the engineers are able to keep track of the changes over time.
This phase allows data engineers to perform the appropriate data transformations faster and be able to monitor, maintain and scale the data ingestion jobs. Moreover, data engineers can early detect any failed harvesting job and receive informative error logs before data are used in the platform.

Table 2-9: Data Engineer Journey Phase 1 – XAI Preparation

Phase 1: XAI Preparation - Collaborate on uploading / handling data	
Actions	Develop automatic ingestion pipelines through APIs
	Ensure consistent data understanding during ingestion via data mapping
	Define & implement data anonymisation operations (on premise)
Challenges Addressed	Able to design and implement data ingestion jobs for XAI pipelines: ensure data quality and conformance to a common data model for data with temporality through an API harvesting schedule

Phase 2: XAI Experimentation - Collaborate on the design of XAI pipelines

In this phase, data engineers need to make sure that designing XAI pipelines will involve all the necessary parts for the procedure to be easily implemented, controlled and scaled in production. This involves the collaboration with the data scientists during the design of a pipeline. Designing a visual workflow is of great assistance to both roles. The XMANAI platform also provides reusable pipelines (e.g., for feature transformations and ML/DL), where each one can be tested independently for scaling and performance. The experimental tracking environment is another tool for mutual collaboration. Both data engineers and data scientists can view the experimental results and assess the relevance of features and the importance of parameters in the model's performance. Each role can then provide feedback in the form of comments or annotations. This collaboration definitely reduces the time required for AI pipeline design, but also improves portability, scaling and performance for the tested pipeline making it ready to move to production.
Data engineers, however, need to handle some challenges in this phase as well. They need to ensure that data scientists do not build hard-coded pipelines that cannot be used in production settings and that the pipelines are portable and compatible with all architectures. In addition, recognizing the edge cases of data pipelines and testing them before deployment is a demanding process that needs to be done, while ensuring data scalability at all stages of the pipeline.

Table 2-10: Data Engineer Journey Phase 2 – XAI Experimentation

Phase 2: XAI Experimentation - Collaborate on the design of XAI pipelines	
Actions	Support data scientists in building production - ready pipelines
	Define metrics to monitor the impact of features on model performance & fairness
	Improve (re-design if needed) the implementation of the data transformation and feature computation tasks to be more scalable
	Test and debug the pipeline, including the generated models
	Integrate model risk assessment & model documentation and explainability procedures in the pipeline design and implementation
Challenges Addressed	Ensure pipeline is portable and compatible with any architecture through the view of available pipelines and their details

	Ensure data scientists do not hard-code pipeline logic that cannot be used in a production setting through collaboration
	Testing data pipelines for edge cases before deploying them in production
	Ensure scalability of data in all steps of the pipeline

Phase 3: XAI Application - Deploy XAI pipelines to production

Phase 3 includes all the processes required for the XAI pipeline to be efficiently deployed to production and be monitored in real-life manufacturing settings. At runtime, the XMANAI manufacturing apps are expected to leverage the deployed containerized pipelines by triggering the execution of a pipeline that may occur based on schedule or on-demand. To this end, data engineers can access the XMANAI platform to monitor and handle the execution and orchestration of all the processes of a pipeline and make sure that the executions work as planned.
Through the execution logs, they know what happened with each execution of the pipeline so as to troubleshoot any failures and identify drifts and errors in the data produced. In parallel, they are able to monitor the system performance metrics and request for scaling the available resources. The data engineers are also responsible to ensure that pipeline outputs are leveraged by the correct manufacturing application, each one containerized properly.

Table 2-11: Data Engineer Journey Phase 3 – XAI Application

Phase 3: XAI Application - Deploy XAI pipelines to production	
Actions	Create containerized AI-enabled applications
	Automate development of validation, training, tuning, and inference in AI pipelines
	Implement model serving processes
	Monitor & tune model performance, fairness and bias in production
	Define and monitor constraints, metrics and distributions on inputs and outputs of AI pipelines
	Monitor drift between training environment and reality
	Deploy new pipeline versions and retrained models as they become available
Challenges Addressed	Spiky & intense workloads (common in AI pipelines) make resource planning difficult, but this can be overcome with continuous monitoring of resources and performance metrics and proper testing before deployment
	Slowly induced bias and drift causing models to deteriorate can be detected through execution logs
	Ensure comprehensive observability of the developed pipeline by all involved teams - ensure the metrics that are monitored and the way various results are visualised are actionable

3 Technical Requirements

This section introduces an updated version on the technical requirements that were presented in D1.3, as well as the description of their acceptance criteria and success and failure results.

3.1 Overview

A requirement is a service, function or feature that a user needs in the software. Requirements can be functions, constraints, business rules or other elements that must be present to meet the need of the intended users. Requirement gathering techniques in agile software development come in many shapes and forms, but the most common form is a User Story. A User Story is a requirement expressed from the perspective of an end-user goal. User stories represent the needs of the customer in a simply written narrative that can be easily understood.

The main difference between user stories and use cases (as another common technique of requirement elicitation) is their objectives. The user story focuses on the experience and what the person using the product wants to be able to do. A traditional use case focuses on functionality and what the product should do.

One of the principles behind User Stories is that the product could be fully represented through the needs of its users. Because the User stories are short and simple descriptions of a feature told from the perspective of the person who desires the new capability, usually a user or customer of the system. The focus is on why and how the user interacts with the software. A user story is essentially a high-level definition of what the software should be capable of doing.

User story descriptions typically follow a simple template as a Card:

As a <role>, I want <goal> so that <benefit>

Bill Wake (2003) proposed the INVEST acronym that expresses the six characteristics which a proper user story should have.

- *Independent* – One user story should be independent of another (as much as possible). Dependencies between stories make planning, prioritization, and estimation much more difficult.
- *Negotiable* – Details of the story can be worked out during an Iteration planning meeting. A story with too much detail can limit conversations (at times).
- *Valuable* – Value to the customer needs to be evident.
- *Estimable* – There needs to be enough detail for the developers to estimate a user story to allow prioritization and planning of the story.
- *Small* – A good story should be small in effort, typically no more than 2-3 person weeks of effort.
- *Testable* – User stories should be testable with certain acceptance criteria.

To ensure that the XMANAI Platform meets the requirements of the different stakeholders (described in Section 2.2-2.4), an agile methodology is implemented for requirements elicitation, promoting interactive sessions with the XMANAI partners. Requirements of the XMANAI platform are extracted in the form of user stories, collaboratively by technical partners and business stakeholders through interactive sessions. XMANAI has been using Miro (www.miro.com), which is an online collaborative whiteboard platform, configured with separate User Story boards for each of the roles in XMANAI, namely Data Scientist, Data Engineer and Business Expert.

The interaction between the partners to fill-in the User Story boards has stimulated creativity and provided the opportunity to examine the needs not only from a technical point of view but also from a practical point of view, defining the behaviour of the XMANAI platform towards the initial goal of achieving explainability in AI for manufacturing purposes. Several user stories have been extracted at the end of the brainstorming sessions. Later, the extracted user stories were tagged and categorized by technical WPs of the XMANAI project, namely WP2, WP3 and WP4, and their corresponding tasks in order to be consolidated in the XMANAI requirements backlog.

Figure 3-1 represents the User Story board that was created collaboratively by the XMANAI partners for Data Scientists.

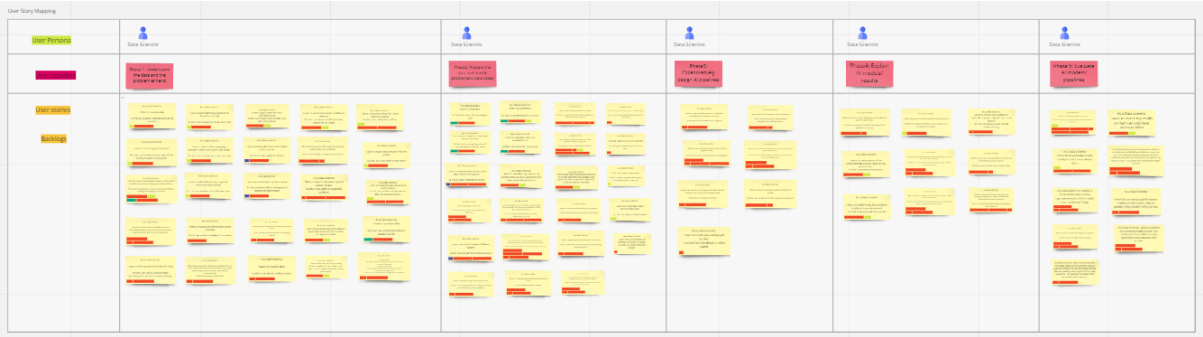

Figure 3-1: Miro User story board for Data Scientist

Figure 3-2 presents the User Story board that was created collaboratively by the XMANAI partners for Data Engineers.

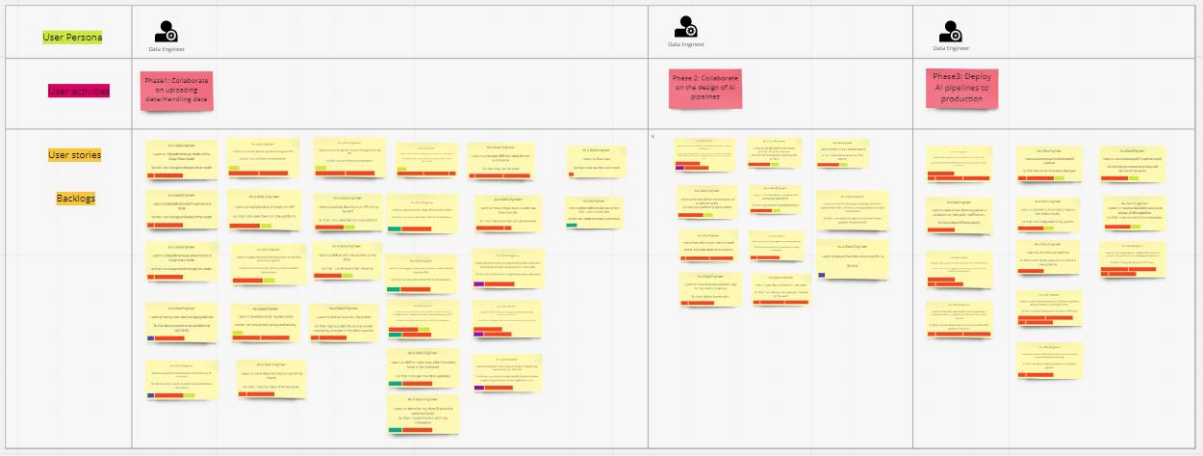

Figure 3-2: Miro User story board for Data Engineer

Figure 3-3 presents the User Story board that was created collaboratively by the XMANAI partners for Business Experts.

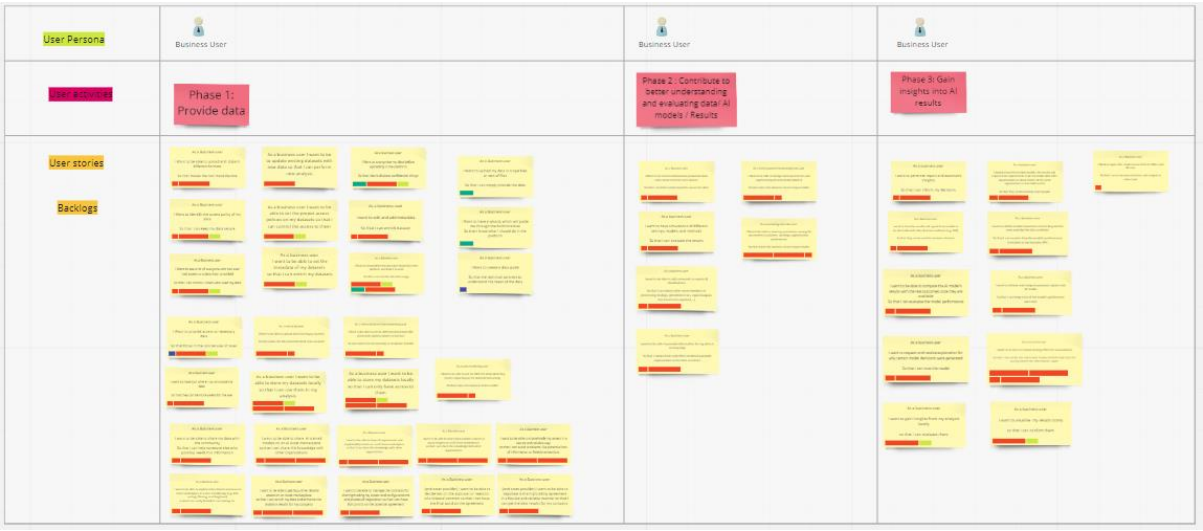

Figure 3-3: Miro User story board for Business User

It is also useful to define non-functional requirements that can bring additional quality characteristics to the user stories. They serve as constraints that can apply to a set of functional requirements and allow the user to understand the attributes of the requirements rather than its functional behaviours. The non-functional requirements define attributes such as: availability; maintainability; performance; reliability; scalability; security and usability.

Currently exists three common ways to define non-functional requirements:

- **Backlog Item** – Independent requirement as an individual item that normally applies if the requirement is complex and difficult to implement.
- **Acceptance Criteria** – Is allocated to the respective requirements with the definition of conditions to be accepted, normally used when applied to backlog items individually.
- **Definition of Done** – In case of the requirement being applicable to the entire solution.

With the requirements, both technical and non-functional, properly defined, an additional step was taken; for each requirement, a categorization of priority was defined to better comprehend which requirements are more important to ensure the XMANAI MVP.

3.2 Backlog

This section focuses on the backlog of the XMANAI platform and provides a detailed presentation of all technical requirements in the form of user stories. The XMANAI platform backlog is extracted collaboratively through brainstorming by partners through Miro. After several interactive sessions and further discussion, the XMANAI technical requirements have been updated. A high-level overview of the technical requirements including user stories descriptions was provided in D1.2 and D1.3. Table 3-1 in this deliverable, provides the list of the technical requirements after their revision from D1.3. In addition to the description of user stories, the system requirements through the description of acceptance criteria and success and failure results and the priority level regarding the MVP in a scale of low, medium and high priority, have been added. The updates in respect to the previous release are also clearly highlighted.

Table 3-1: Final XMANAI Technical Requirements

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_1	T2.1	H	Data Scientist, Data Engineer, Business User	add/edit/remove new data sources for data/metadata import	Data/metadata from these sources can be available for analysis	Any user can add a new data source. New data sources are stored inside Pipe Objects, which are treated as assets. Editing or removing of a data source equals the manipulation of an asset and will be handled according to the set policies.	New data source has been created/an existing data source has been edited/an existing data source has been deleted.	Error shown	Added Priority, Acceptance Criteria, Success and Failure
TR_2	T2.1	M	Data Scientist, Data Engineer, Business User	define rules how often the data have to be collected	I can get regular data updates from an external source	If the user is creating a new data source, he can schedule the regular updates as he pleases. In case he wants to (re-)schedule an existing data source, which was not created by him, he needs the rights to do so set in the policies of the Pipe Object of the corresponding data source.	The automatic update cycle of a dataset yielded from an external data source has been set.	Error shown	Added Priority, Acceptance Criteria, Success and Failure
TR_3	T2.1	H	Data Scientist, Data Engineer, Business User	upload data as single file or batch of files	I can work with these data in the platform	user can upload a file in binary format. The access policy is set accordingly	File is added to the Assets Store, policy is created in the Policy Manager	File is not created, no access policy is set and metadata is not stored.	Added Priority, Acceptance Criteria, Success and Failure
TR_4	T2.1	H	Data Engineer	Delete data and its associated metadata	I can remove data, which I don't need anymore	The user who has created the data is only allowed to delete the data. The metadata is retained in the metadata manager.	The data is deleted, metadata is not deleted.	The Asset is retained in the database.	Added Priority, Acceptance Criteria, Success and Failure

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_5	T2.1 & T3.1	M	Admin	manually create / import a specific domain (Category) for the data model	relevant data can be mapped to it and become available in the platform	The admin is capable of add manually a new domain in the data model	Domain is created and is possible to navigate to it	Proper error shown	Added Priority
TR_6	T2.1 & T3.1	M	Data Scientist, Business User	explore the data model to find concepts, fields and relationships	I can ensure my specific data needs are addressed	The user should be capable of navigate through the data model, analyse the characteristics of each node and is relationships with the rest	Graph visualization of the data model is presented to the user	Proper error shown	Added Priority
TR_7	T2.1 & T3.1	H	Admin, Data Engineer	manually add new concepts, properties and relationships to the data model	I can ensure the needs of data scientists and business users are addressed	The user requests the addition of a concept, property or relationship to the data model	Admin approves the request and is added to the data model	Admin rejects the proposal for the new addition	Added Priority
TR_8	T2.1 & T3.1	H	Admin, Data Engineer	manually update the concepts, fields and relationships on the data model	I can change the model over time	User requests an update to an existing concept, property or relationship on the data model	Admin approves the request and the characteristics of the respective node are updated	Admin rejects the proposal for the update	Added Priority
TR_9	T2.1 & T3.1	L	Admin, Data Engineer	manually control the different versions of the data model	I can view and retrieve all the history of edits	User can request for a log of the modifications made to the data model	Retrieved a log of the modifications	Proper error shown	Added Priority
TR_10	T2.1 & T3.1	M	Admin, Data Scientist	view the data model	I can use the model per domain (category)	User can visualize only a part of the data model, filter by domain	Graph visualization of the domain with respective nodes related	Proper error shown	Added Priority
TR_11	T2.1 & T3.1	L	Data Scientist	generate different representations (e.g. graph) of the data model	I can make it available to other applications / components	User requests the data model in text format	Users visualize the data model in a descriptive form	Proper error shown	Added Priority

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_12	T2.1	H	Data Scientist, Data Engineer, Business User	add/edit metadata for my data (based on a common metadata model)	I can improve the quality of data for further reuse	Some basic metadata of assets such as the name and UUID is created automatically along with the asset in the database.	Metadata is added to the Metadata Manager	No metadata is available for the asset.	Added Priority, Acceptance Criteria, Success and Failure
TR_13	T2.1	H	Data Scientist, Data Engineer, Business User	define how my data are mapped to a data model	the data types and semantics of my data can be available to anyone who uses my data	File Data Harvester: According to the set policies the user has to have the rights in order to access and edit the file, which he wants to convert to the XMANAI data model.	File Data Harvester: The conversion was successful, and the asset can now be stored inside the Asset store	Error shown	Added Priority, Acceptance Criteria, Success and Failure
TR_14	T2.1	Deprecated	Data Scientist, Data Engineer, Business User	define whether and how my data should be cleaned	I can increase their quality before they are stored	-			Deprecated requirement
TR_15	T2.1	H	Data Scientist, Data Engineer, Business User	manage (create new, rename, move, delete) my datasets	I can work with them in the platform	The asset owner has the rights to manage (create new, rename, move, delete) his/her assets. When a user creates an asset, the access policy is presented to the user to modify the visibility of this asset.	A new dataset is created or an existing dataset is modified.	No new dataset is created or the existing dataset is not updated.	Added Priority, Acceptance Criteria, Success and Failure
TR_16	T2.1	H	Business user, Data Scientist, Data Engineer	update existing datasets with new data	I can perform a new analysis	The asset owner can update an existing dataset through the API Data Harvester by providing the pipeID of the asset.	An existing dataset is updated.	The data is not updated, and the older version is retained in the database.	Added Priority, Acceptance Criteria, Success and Failure

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_17	T2.1 & T2.3	H	Data Scientist	download data that I have legitimate access as a file	I can use them offline	If the user has the proper access rights, the datasets in tabular format and binary files can be downloaded.	A binary file is downloaded	Error shown	Added Priority, Acceptance Criteria, Success and Failure
TR_18	T2.1 & T2.3, T5.X	H	Data Scientist	retrieve data that I have legitimate access through an API	I can use them in the XMANAI manufacturing apps	If the user has the proper access rights, the data can be retrieved from the database API	Depending on the type of the data, a binary file or a json output is received.	Error shown if there are insufficient access rights or the requested data does not exist in the database	Added Priority, Acceptance Criteria, Success and Failure
TR_19	T2.1 & T2.3, T3.2, T3.3	M	Data Scientist, Data Engineer	export data samples from the dataset	I can view them in other XMANAI components and external tools	If the user has the proper access rights, certain fields of a tabular asset can be retrieved using the database API	Certain fields of the dataset, not the whole dataset is received as output	Error shown if there are insufficient access rights or the requested data does not exist in the database	Added Priority, Acceptance Criteria, Success and Failure
TR_20	T2.1, T2.2, T3.5	H	Business user, Data Engineer	be able to store my datasets on my premises	I can only have access to them and I can use them in my analysis	The asset owner is able to store his/her asset on the on-premise environment of the platform	The asset owner stores the assets of his/her organisation on the on-premise environment of the platform	The asset owner cannot store the assets of his/her organisation on the on-premise environment of the platform	Added Priority

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_21	T2.2	H	Business user, Data Scientist	able to define the proper access policies on my assets (datasets, AI models, analysis results)	I can define who shall have access to my assets and under which circumstances	The asset owner can set the proper access policies that define which organisations can view and potentially request access to their assets through the marketplace	Access policies can be defined for the assets and only the eligible organisations can view the assets and potentially request access to the assets through the marketplace	Access policies cannot be defined for an asset or the asset can be viewed despite the access policies set.	Added Priority, Adapted the Description as AI pipelines and Experiments are assigned by default (non-configurable by the user) organisation-based access policies from the XMANAI platform.
TR_22	T2.2	H	Business user, Data Scientist	able to combine multiple access policies on my assets (datasets, AI models, analysis results)	I can define more complex access restrictions to my assets	The asset owner is capable of setting and combining multiple access policies that define which organisations can view and potentially request access to their assets through the marketplace	Access policies can be defined and be combined for the assets and only the eligible organisations can view the assets and potentially request access to the assets through the marketplace	Access policies cannot be defined and combined for an asset or the asset can be viewed despite the access policies set.	Added Priority, Adapted the Description as AI pipelines and Experiments are assigned by default (non-configurable by the user) organisation-based access policies from the XMANAI platform.

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_23	T2.2	H	Business user, Data Scientist	able to define access policies based on various attributes of the requestor (datasets, AI models, analysis results)	I can define who shall have access to my assets and under which circumstances	The asset owner can set the proper access policies by selecting specific attributes of the organisations that can view and potentially request access to their assets through the marketplace	Access policies can be defined for the assets based on the various attributes of the organisations and only the eligible organisations can view the assets and potentially request access to the assets through the marketplace	Access policies cannot be defined for an asset based on the various attributes of the organisations or the asset can be viewed despite the access policies set.	Added Priority, Adapted the Description as access policies are based on requestor's attributes. AI pipelines and Experiments are assigned by default (non-configurable by the user) organisation-based access policies from the XMANAI platform.
TR_24	T2.2	H	Business user, Data Scientist	able to update or remove the access policies on my assets (datasets, AI models, analysis results)	I can reconsider who shall have access to my assets	The asset owner can update or remove applied access policies at any point	Access policies can be modified or deleted at any point with immediate effect	Access policies cannot be modified or deleted by the asset owner	Added Priority, Adapted the Description as AI pipelines and Experiments are assigned by default (non-configurable by the user) organisation-based access policies from the XMANAI platform.

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_25	T2.2	H	Business user, Data Scientist	able to define the access level of my assets only to my organisation	I can provide access only to my organisation's users	The asset owner can restrict the access to the asset only to his/her organisation	Access can be restricted to organisation of the asset owner	Access cannot be restricted to organisation of the asset owner	Added Priority
TR_26	T2.2	Deprecated	Business user, Data Scientist	able to define the access level of my assets only to selected users outside my organisation	I can get support from other data scientists	-	-	-	Deprecated requirement
TR_27	T2.2	H	Business user, Data Scientist	able enforce the access control decision based on my access policies	I can ensure that the access to my assets is always safeguarded	The access to every asset is regulated through the access policies set by the asset owner	The access to each asset is regulated by the access policies as set by the asset owner	Access to each asset is not regulated by the access policies set by the asset owner	Added Priority
TR_28	T2.2	H	Data Scientist, Data Engineer, Business User	ensure that only properly authenticated users have access to my assets (datasets, AI models, AI pipelines, experiments, analysis results)	I can ensure their privacy and security	Access to the assets of the platform is only possible for authenticated users if they are eligible to view them based on the access policies set by the asset owner	Only registered user can view and request access to assets that they are eligible to view and access based on the access policies set by the asset owner	Unregistered users can view or request access to assets of the platform	Added Priority
TR_29	T2.2	M	Data Scientist, Data Engineer, Business User	configure and apply different data anonymisation processes on my data before they are uploaded in the platform	I can assure no personal or critical information is disclosed	The assets owner can apply different anonymisation processes to his/her assets prior to making them available in the platform	Asset owners can define and execute anonymisation processes to their asset prior to uploading them in the platform	No anonymisation process can be defined and executed prior to uploading the asset	Added Priority

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_30	T2.2	M	Data Scientist, Data Engineer, Business User	configure and apply different data anonymisation processes on samples of data before they are uploaded in the platform	I can understand what changes are to be performed on my data	The assets owner can apply different anonymisation processes on samples of the assets prior to making them available in the platform	Asset owners can define and execute anonymisation processes on samples of their assets prior to uploading them in the platform	No anonymisation process can be defined and executed on samples of the assets prior to uploading the asset	Added Priority
TR_31	T2.2, T5.x	L	Business user, Data Scientist	ensure that my data are transferred between the different layers of the platform securely	I can ensure that my data will not be disclosed to unauthorized parties	The data in transit are securely transferred between the different layers of the platform	The proper security measures are applied on data in transit between the different layers	The data in transit are not properly safeguarded when they are transferred between the layers	Added Priority
TR_32	T2.3	M	Business User, Data Scientist, Data Engineer	share my assets (e.g. datasets, AI models, features, analysis results) with other organisations / users of my preference	I can help someone else who possibly needs this information	Given the data provider wants to share an asset, When the provider defines the metadata for the access policies of the asset and marks the asset as shareable, Then the asset becomes available in the marketplace.	Asset can be viewed inside the marketplace by the users that have appropriate access rights to view it.	Asset remains private and can be viewed only by users of the same organisation as the data owner/provider	Added Priority
TR_33	T2.3	Deprecated	Business User	trade my assets (e.g. datasets, AI models, features, analysis results) in a secure and reliable way for a specific time period	I can gain new revenues from my assets	-	-	-	Deprecated requirement

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_34	T2.3	M	Business User, Data Scientist, Data Engineer	search and explore other data/AI assets on an in a user-friendly way (e.g. based on metadata, with sorting, filtering, matching level)	I can easily find what I am looking for	Given a search keyword term (or a sorting/filtering command), When the search button is pressed, then results (assets) appear in a list view based on the search query criteria and user rights (to view datasets).	The results are correct and verified based on the search query criteria	Results are not the ones anticipated based on the search query criteria	Added Priority
TR_35	T2.3	M	Data Scientist, Data Engineer	view metadata of the selected asset (e.g. datasets, AI models, AI pipelines, analysis results)	I can determine if the asset addresses my needs	Given the user is authorised to view this asset in the catalogue, When the asset is selected, then a view with all the asset details (metadata) is presented.	The user can locate the asset and view the details (metadata) of the asset	The asset does not appear at all in the catalogue or its metadata cannot be loaded	Added Priority
TR_36	T2.3 & T2.4	M	Business User, Data Scientist, Data Engineer	be informed about the derivation procedure for the asset and the associated rights/licences	I can take an informed decision for the asset acquisition	Given the user is authorised to view this asset in the catalogue, when the asset is selected a view with all the asset details (metadata), including associated rights/licenses and how the asset was derived(stages) is presented.	The user can locate the asset and view the derivation process and license details of the asset	The asset's license or derivation info does not appear at all in the catalogue/asset's page	Added Priority

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_37	T2.3	M	Business User, Data Scientist, Data Engineer	get access to assets (e.g. datasets, AI models, analysis results) created by other users, in a secure and reliable way for a specific time period	I can enrich my data and enhance the analytics results for my company	Given a data consumer desires to get access to an asset he located in the marketplace he makes a request to get access, then the data owner is informed, and he can form a new data sharing agreement, which may have an expiration date in its terms and send it to the consumer. Given the data consumer agrees to the terms when the data owner agrees as well, then the contract becomes valid, and the consumer has access to the asset in a secure and reliable manner.	Data consumer has access to the asset for a period of time set in the contract's terms.	Data consumer does not have access to the asset at all.	Added Priority
TR_38	T2.3	Deprecated	Business User, Data Scientist, Data Engineer	buy assets (e.g. datasets, AI models, features, AI pipelines, experiments, analysis results) created by other users, in a secure and reliable way for a specific time period	I can enrich my data and enhance the analytics results for my company	-	-	-	Deprecated requirement

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_39	T2.3	M	Business User, Data Scientist, Data Engineer	manage the contracts for sharing my assets and configure contract terms	I can have full control on the potential agreement	Given an authorized user wants to create a contract, when he navigates to the corresponding asset, Then he can create a new contract/ sharing agreement. Given an authorized user wants to manage a contract, when he navigates to the contracts list, then he can view the open/ongoing contracts and define/update their terms, approve or reject them.	User can create/update the terms and details of a contract when it is not yet valid.	User cannot create/update a contract and its terms. Failure notification shown.	Added Priority
TR_40	T2.3	M	Business user, Data Scientist, Data Engineer (and asset provider)	be able to make negotiations over the terms of a sharing agreement	I can achieve the best outcome for my company	Given an asset contract negotiation process has commenced, when either party desires to propose new set of terms, Then the authorized user can do so through the contract page.	User can update terms and the other party is informed about it automatically to approve or reject them.	User cannot update terms and is unable to move the negotiation process forward.	Added Priority
TR_41	T2.1 T2.3, T3.2	M	Data Scientist	be able to locate relevant datasets to my own	I can enrich my data and fill-up missing information	Given a data consumer seeks datasets with specific characteristics, when he navigates to marketplace, then he can search the catalogue for relevant datasets	User finds relevant datasets that he can use to enrich his own data.	User gets no results that can be deemed relevant to his own data.	Added Priority
TR_42	T2.4	M	Data Scientist, Data Engineer	view how often the data is updated and when	I can know if I have the latest data or change the update frequency if needed	Only the asset owner with proper access rights can view an overview of data updates and decide the update frequency	The overview can be visualized	Error shown if there are insufficient access rights, or the requested data does not exist in the database	Added Priority, Acceptance Criteria, Success, Failure

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_43	T2.4	M	Data Scientist	have a control version of the datasets	I am sure that I use the latest updated data	Users with access rights can create and retrieve newer and older versions of the data.	The latest updated version of a dataset is available.	Error shown	Added Priority, Acceptance Criteria, Success, Failure
TR_44	T2.4 & T3.3, T4.1, T4.2	M	Data Scientist, Data Engineer	have a control version of assets (AI models, AI pipelines, features, experiments, results)	I can keep track of the changes introduced and limit the impact of changes on existing pipelines	If the metadata and name of the asset can be kept constant for all the versions, multiple versions of a binary and tabular asset belonging to all asset types can be stored and retrieved.	Several versions of an asset can be retrieved and created	Error shown	Added Priority, Acceptance Criteria, Success, Failure
TR_45	T2.4	M	Business user	XMANAI to register each access event (based on actions performed) of other users to my data or other assets	I can have detailed logs who and when accessed my data and other assets	Users with sufficient access policies can view the event log	The event log is displayed	Error shown if there are insufficient access rights, or the requested data does not exist.	Added Priority, Acceptance Criteria, Success, Failure
TR_46	T2.2 & T2.4	M	Business user	see a list of users who has ever had access to an asset that I provided and what activities were performed	I can monitor my assets usage	The list of users that have obtained access to my assets is maintained and presented to the asset owner	The asset owner can see the list of users that have obtained access to his/her assets	The asset owner cannot see which users have obtained access to his/her assets	Added Priority
TR_47	T2.4	M	Business user	view which data and assets I have shared with whom	I can monitor my sharing activities	The list of organizations that have obtained access to my assets through an active sharing contract is maintained and presented to the asset owner	The asset owner can see the list of organizations that have obtained access to each asset	The asset owner cannot see the list of organizations that have obtained access to each asset	Added Priority

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_48	T2.2 & T2.4	M	Business user	XMANAI to register to whom I permitted access to my data or other assets	I can log the sharing my data and other assets with other users	The list of users that have obtained access to my assets is maintained and presented to the asset owner	The asset owner can see the list of users that have obtained access to his/her assets	The asset owner cannot see which users have obtained access to his/her assets	Added Priority
TR_49	T2.4	M	Data Scientist	check the IPR of the assets involved in an AI pipeline	all assets included in a pipeline are used in alignment with their licenses	The user can view and select assets in an AI pipeline that he/she is authorised to use/	The user can use an asset in accordance with its IPR (active contract, read/write access) in a pipeline	The user can use all or no assets irrespectively of the active contracts while designing a pipeline	Added Priority, Acceptance Criteria, Success, Failure
TR_50	T3.2	M	Data Scientist	query data to which I have legitimate access	I can find a subset of the data I can use in my analysis	Given an authorized user able to query data (i.e. belongs to organisation that is data owner/provider), when the user navigates to the data preparation view he can use the functionality to query/filter data and isolate a subset of data.	User can query/filter data	Unable to view data or query data, if unauthorized.	Added Priority
TR_51	T2.1 & T3.2	M	Data Scientist	be informed about the data types and semantics per feature that exists in my dataset	I can quickly understand the data that I will use in an analysis	The user can see the data type of all features that appear in a dataset whenever necessary.	The user can view the data types while using the pipeline designer.	The user cannot access any info for the dataset while using the pipeline designer.	Added Priority, Acceptance Criteria, Success, Failure
TR_52	T3.2	M	Data Scientist	preview a sample of the data	I can obtain a more concrete understanding of the data at hand	Given the user is authorised to use the specific dataset, when he executes a data preparation task, then a sample of data is retrieved to be shown to the user in the logs.	The user views a sample of the data	Error shown	Added Priority. Adapted the Acceptance Criterion for more clarity

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_53	T3.2	M	Data Scientist	view data distribution/profiling charts or summary statistics for the data (e.g. number of missing values, min/max values)	I can monitor data drifting issues	Given the user is authorised to use the specific dataset, when he selects to view summary statistics, then a set of metrics or informative charts is shown to the user.	The user views the summary statistics and useful charts of the data.	Error shown	Added Priority
TR_54	T3.2	Deprecated	Data Scientist	define rules for input data	I quickly identify wrong values caused by changes in the data sources	-	-	-	Deprecated requirement
TR_55	T3.2	M	Data Scientist	create new features based on the current data (like min, max, mean values) that will be part of the same dataset	I can have more informative datasets, depending on the task	Given a working dataset with the appropriate features' datatypes, when a user wants to derive new features from current ones, Then he selects the functionality and applies it to the dataset	New features are created as part of the same dataset	Error shown	Added Priority
TR_56	T3.2	M	Data Scientist	handle missing values (impute)	I can prepare the data for the subsequent AI analysis	Given a working dataset with the appropriate features' datatypes, when a user wants to impute missing values, Then he selects this functionality and applies it to the dataset	Missing values are filled-in according to the conditions and rules selected.	Error shown	Added Priority
TR_57	T3.2, T3.3	M	Data Scientist	encode categorical data	I can prepare the data for the subsequent AI analysis	Given a working dataset with the appropriate features' datatypes, when a user wants to encode categorical values, then he selects this functionality and applies it to the dataset	Categorical data are encoded in a numeric format.	Error shown	Added Priority. Updated Tasks.

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_58	T3.2, T3.3	M	Data Scientist	apply scaling and data normalization	I can prepare the data for the subsequent AI analysis	Given a working dataset with the appropriate features' datatypes, when a user wants to scale/normalize values, Then he selects this functionality and applies it to the dataset	Data are scaled/normalised according to the selected parameters.	Error shown	Added Priority. Updated Tasks.
TR_59	T3.2	M	Data Scientist	easily split the data for training and evaluation (data segmentation)	I can train and apply the models as I see fit	Given a working dataset with the appropriate features' datatypes, when a user wants to split the data into a training and testing set, then he selects this functionality and applies it to the dataset	Data are split into training and testing (for evaluation) sets	Error shown	Added Priority
TR_60	T3.2	M	Data Scientist	apply simple transformations on the data	I make them more appropriate for processing and visualisation	Given a working dataset with the appropriate features' datatypes, when a user wants to perform transformations, then he selects the appropriate functionality and applies it to the dataset	Data are transformed according to the selections of the user.	Error shown	Added Priority
TR_61	T3.2	M	Data Scientist	change the data type of some features	I can manipulate them according to the needs of an AI model (e.g. convert to datetime)	Given a working dataset, when a user wants to modify the data type of a feature, then he selects the appropriate functionality and applies it to the selected features.	Data types for specific features are modified based on user selection.	Error shown	Added Priority
TR_62	T3.2	L	Data Scientist	apply data augmentation techniques	I can reduce overfitting of the models	Given a working dataset, when a user wants to augment data, then he selects the appropriate functionality and applies it to the dataset.	The amount of data is increased based on the selections and parameterization of the user	Error shown	Added Priority

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_63	T3.2	M	Business User	perform calculations over my data	I can keep track of important Key Performance Indicators that are important for my business	Given a working dataset, When a user wants to perform calculations over the data, then he selects the appropriate functionality and applies it to the selected features	Calculations are performed on the dataset	Error shown	Added Priority
TR_64	T3.2 (& T3.1)	M	Business User	be able to provide information for my data in an easy way	I reduce time and effort needed to provide explanations to the data scientists	Given a user uses a dataset, when he wants to find information regarding its structure and meaning, then he views the relevant information from the data model.	Informative details for the structure and semantics of the asset are retrieved and displayed	Authorisation error shown	Added Priority
TR_65	T3.3	M	Data Scientist, Data Engineer	define and configure an AI pipeline for training, testing and/or production purposes	I can provide a solution for a specific problem	Given a user wants to define and configure an AI pipeline, when he navigates to the Pipeline Designer, then he can design and configure an AI pipeline.	An AI pipeline is defined and configured for training, testing and/or production purposes.	Error in the configuration or storage of the AI pipeline is shown	Added Priority
TR_66	T3.3 (& T4.1 / T4.2)	M	Data Scientist, Data Engineer	include compatible baseline algorithms in an AI pipeline	I can provide a solution for a specific problem	Given a user wants to use a baseline model in an AI pipeline, when he navigates to the Pipeline Designer, Then he can access and configure the parameters of such a model.	An AI pipeline includes appropriate baseline models compatible for a specific problem	Appropriate error is shown	Added Priority. Adapted the Acceptance Criterion for more clarity
TR_67	T3.3 (& T4.1 / T4.2)	M	Data Scientist, Data Engineer	include compatible trained models in an AI pipeline	I can provide a solution for a specific problem	Given a user wants to add an already trained (pre-trained) model to an AI pipeline, when he navigates to the model selection list, Then he can select his model and add it to the pipeline.	An AI pipeline includes appropriate models trained for a specific problem	Appropriate error is shown	Added Priority. Adapted the Acceptance Criterion for more clarity

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_68	T3.3	M	Data Scientist	register a trained AI model I have created as part of an AI pipeline	I can reuse it in my AI pipelines	Given a user wants to add an already trained (pre-trained) model to an AI pipeline, when he navigates to the model selection list, then he can select his model and add it to the pipeline.	The trained model is added to the pipeline successfully.	Model not found or model is not added to the pipeline successfully (error thrown)	Added Priority.
TR_69	T3.3, T2.3	M	Data Scientist, Data Engineer	collaborate in the configuration of AI pipelines with selected users (within my organization)	I can create optimal workflows for a specific problem	Given a user wants to collaborate with other users of the same organisation during AI pipeline configuration, when he starts a new pipeline, then automatically other users of the same organisation can view it, contribute and provide comments/feedback.	Collaboration in the configuration of AI pipelines with selected users of the same organisation is available. They can post comments in the pipeline or its tasks.	Users cannot view or edit a pipeline created by other users in their organization.	Added Priority, Failure.
TR_70	T3.3	Deprecated	Data Scientist, Data Engineer	define pipelines that can be used as templates for specific problems	my colleagues and I can reuse them	Given a user wants to create various AI pipelines, when he saves a pipeline, then he may opt to save it as a template for specific problems. Given a user wants to reuse a template AI pipeline, when he creates a pipeline, then he may use a template for specific problems.	AI pipeline templates are available to be reused.	Error shown	Deprecated requirement
TR_71	T3.3	Deprecated	Data Scientist, Data Engineer	clone a designed pipeline	I can create alternative version of the pipeline and improve its performance without re-creating it from scratch	Given a user wants to reuse an AI pipeline configuration, when he creates a new pipeline, then he may opt to clone an existing one (instead of starting from scratch) and make the appropriate adjustments.	Duplicate/clone pipeline is created and can be edited.	Error shown	Deprecated requirement

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_72	T3.3	Deprecated, TR_70	Data Scientist, Data Engineer	join designed pipelines	I can create advanced combinations (e.g. training pipeline with testing pipeline, multiple training pipelines)				Deprecated requirement
TR_73	T3.3	Deprecated	Data Scientist, Data Engineer	execute step-by-step an AI pipeline over sample data	I can ensure that the result is the intended one	Given a user wants to test an AI pipeline, when he feels ready with the design, then he can select to execute it up to the point he desires over sample data.	A test run is executed, step-by-step, up to the selected point of the AI pipeline.	Error shown	Deprecated requirement
TR_74	T3.3	Deprecated	Data Scientist, Data Engineer	reuse common features in different pipelines	I do not need to recompute them	-	-	-	Deprecated requirement
TR_75	T3.3 (& T4.1 & T4.2)	M	Data Scientist	configure training to control parameters, such as learning rate reduction when a metric has stopped improving or stop it	I can notify if there is any problem during training	Given a user wants to configure the training of a new model, when he has chosen the baseline model, then he can configure different the training control options allowed	Notification when trigger conditions are present		Added Priority
TR_76	T3.3	Deprecated	Data Scientist	receive recommendations for automated feature selection	I can perform feature engineering in a faster / easier manner, in cases of high-dimensionality data	-	-	-	Deprecated requirement

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_77	T3.3	H	Data Scientist	choose methods to apply for explaining an AI pipeline (including input data, models, and results)	I can select the ones that best fit with the problem I am solving	An appropriate explainability technique can be selected together with the respective ML/DL model in the pipeline.	The user can include an available explainability technique.	The user cannot bind explainability techniques with the models or pipelines.	Added Priority, Acceptance Criteria, Success, Failure.
TR_78	T3.3	Deprecated	Data Scientist	properly adjust the explanations depending on the user profile	I can increase understanding of the results for its intended users	-	-	-	Deprecated as it's considered as duplicate of TR_84
TR_79	T3.3	M	Data Scientist	define summary statistics to be computed for an AI pipeline or part of it	I can explain the behaviour of the inputs and / or outputs of a pipeline / model	Given a user wants to check the results of an AI pipeline up to a specific point, When the pipeline has been executed, then he can view the summary statistics of the execution up to the selected point.	A test run is executed, step-by-step, up to the selected point of the AI pipeline and summary stats of the results can be displayed.	Error shown	Added Priority
TR_80	T3.3	M	Data Scientist	add annotations and comments in AI pipelines	I can better explain the results	Given a user wants to collaborate with other users of the same organisation during AI pipeline configuration, when he works on the pipeline design, Then he can add comments/annotations in the overall configuration or specific tasks in order for other users of the same organisation to view them.	Comments in the pipeline and its tasks can be added.	Error shown	Added Priority, Updated Acceptance

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_81	T3.3 (T3.4 & T3.5)	L	Data Scientist	define whether the analysis results should be saved as a new dataset or update an existing one	I can re-use the analysis results	Given a user wants to save the results of an AI pipeline, then he can select whether each execution of the pipeline will either save it as a new dataset or update (replace/append) the existing one.	The results are saved either as a new dataset or by updating the existing one based on the selection of the user.	The results are saved by default in a new dataset without allowing the user to select	Added Priority, Failure
TR_82	T3.3 (T3.4 & T3.5)	M	Data Scientist, Data Engineer	export the results of an analysis (AI pipeline)	I can create presentations and analysis in other tools (e.g. Office, BI tools)	The user can select and export the results that he/she is eligible to export	The user selects a result from the list of eligible for exporting results and exports it	The user cannot export a result from the list of eligible for exporting results	Added Priority, Acceptance Criteria, Success, Failure.
TR_83	T3.3	M	Data Scientist	add comments to AI pipelines	I can inform other team members of [interesting findings, identified errors, inputs/outputs that need to be explored]	The user can add comments, reply to comments or resolve comments posted by another user.	The user successfully posts comments in the pipeline or in a specific task of the pipeline	Error shown	Added Priority, Acceptance Criteria, Success, Failure.
TR_84	T3.3	M	Business User	properly visualise the explanations depending on the user profile	I can adapt the information I receive, according to my needs	Given a user wants to visualize an explanation in a more technical way, when visualizing the explanation, then he can opt for a more advanced format from the available ones provided to gain deeper knowledge about the explanation	The data is shown in the selected format	Error shown	Added Priority, Acceptance Criteria, Success, Failure.

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_85	T3.3	M	Data Scientist	define the parameters and metrics of the experiments associated with an AI pipeline	I can track and compare my experiments	Given a user wants to define parameters and metrics for an experiment, when he designs the pipeline, then he can select appropriate parameters and metrics from an available set that will appear in the experimentation tracking view.	Experiments can be compared based on defined parameters and metrics by the user.	Error shown	Added Priority, Failure
TR_86	T3.3 (& T4.1 / T4.2)	Deprecated	Data Scientist	run automatic tests on the registered AI models (when including them in an AI pipeline)	I can quickly check that the trained models are robust and fault-tolerant	Automatic validation of AI models registry	Validation according to specified rules	Error shown	Deprecated requirement as the execution of a model needs to be significantly modified to allow such an automatic validation
TR_87	T3.3 (& T4.3)	M	Data Scientist	support the inclusion of different performance metrics as part of an AI pipeline	I can obtain a better picture of my pipelines' / models' effectiveness according to my needs	Given a user wants to evaluate the performance of an AI pipeline, when he designs the pipeline, then he can select appropriate performance metrics from an available set as part of the pipeline.	Various performance metrics can be selected to evaluate the performance of a pipeline.	Error shown	Added Priority
TR_88	T3.3	M	Data Scientist	generate multiple visualisations as output of an AI pipeline	I can make it available to the involved users	Given a user wants to visualise the results, when he selects the result to be visualised multiple visualisation types are available for selection and generation	The user can select and visualise the results of an analysis through a list of multiple visualisation types	The user is not able to visualise the results	Added Priority

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_89	T3.3	H	Business User	view a visualisation including the results and a link to their explanations	I can take informed decisions	Given a user wants to visualise the results, the selected chart is generated and is accompanied with a link pointing to the explanation of the results	The user can select and visualise the results of an analysis with chart that is supplemented with the proper explanations, for which a user is redirected to the appropriate web page.	The user is not able to visualise the results or the link to the explanation of the results are not made available	Added Priority
TR_90	T3.3	M	Business User	choose among different visualisations	I can create the charts and measurements that help me quickly detect the information I want	Given a user wants to visualise the results, when he selects the result to be visualised multiple visualisation types are available for selection and generation	The user can select and visualise the results of an analysis through a list of multiple visualisation types	The user is not able to visualise the results	Added Priority
TR_91	T3.3	H	Business User	get explanations for why certain predictions were generated in an AI pipeline	I can trust the model	Local explanations are generated for certain predictions if the selected	The user can view explanations for specific predictions only	Error shown	Added priority, Acceptance Criterion, Success, Failure. Adapted Description to better reflect the actual work

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_92	T3.3	Deprecated	Data Scientist	respond to requests for explanations of an AI pipeline generated by users	I can help the business users to trust the model	-	-	-	Deprecated as the requests for explanations are not dynamically created by business users
TR_93	T3.3 (& T2.3)	M	Data Scientist	have a common metadata model for describing assets of the same type (datasets, models, results)	I can share them with other users	Given a user wants to better describe an asset, when he enters the metadata manager and locates the asset, then he can select to add or modify the metadata of the asset based on a common metadata model.	Assets of the same type have a common metadata model so that they can be easily located in the marketplace.	Error shown	Added Priority
TR_94	T3.3 (& T4.3)	M	Business User	define model evaluation metrics (beyond the ones used by the data scientists)	I can monitor how the model's performance translates to my business KPIs	Given a user wants to evaluate the performance of a model, when he tests the model in a pipeline, then he can select appropriate evaluation metrics from an available set.	Model evaluation metrics can be defined.	The selected evaluated metrics are not fulfilled	Added Priority
TR_95	T3.3 (T4.3)	L	Business User	retrieve and compare previous AI model's results with the real outcomes once they are available	I can keep track of the model's performance over time	The values of performance metrics	Three performance metrics are acceptable	Do not meet the requirements of accuracy measurements	Added Priority
TR_96	T3.3	M	Data Scientist, Data Engineer	define where my AI pipeline will be executed	I can ensure the analysis is securely executed in infrastructures under my control	Given a user wants to define where a finalised AI pipeline will be executed, depending on the location in which the pipeline is being scheduled, the appropriate execution environment will be selected	The location of the execution of the AI pipeline is defined	Error shown	Added Priority

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_97	T3.3, T3.4 & T3.5	M	Data Scientist	track the performance of my experiments (AI models) regarding stability and converge of the results, execution times, etc.	I can easily study the available solutions and select the optimal one.	Given a user wants to track the performance of an experiment, when he navigates to the experimentation tracking page, then he can view all experiments along with their characteristics (e.g. execution time, convergence, etc).	The user can have a clear view of all experiments performed along with their performance indicators.	Error sown	Added Priority
TR_98	T3.3, T3.4 & T3.5	Deprecated	Data Scientist	implement different mechanisms to save a checkpoint of the AI model	I can resume the training of a model from the previous point				Deprecated requirement as it's not feasible to be implemented
TR_99	T3.3, T3.4 & T3.5	M	Data Scientist	Have a view of how the hyperparameters performed in each experiment	I can tell when a model is optimized and avoid overfitting	Given a user wants to track the hyperparameter performance of a model, when he navigates to the experimentation tracking page, Then he view all experiments' performance along with their hyperparameter values.	The user can view how models performed based on the hyperparameters selected.	Error shown	Added Priority, Failure
TR_100	T3.3, T3.4 & T3.5	M	Data Scientist	launch and queue the training of different models and with different hyperparameters	I can speed up the experimentation stage	Given a user wants to run experiments with different models and hyperparameters, when selects to execute them, Then the executions are queued to speed up the process.	The different executions are queued to speed up the experimentation process.	Error shown	Added Priority
TR_101	T3.4 & T3.5	M	Data Scientist, Data Engineer	get descriptive error messages	I can debug the AI pipelines that have been executed	The user can see the logs page for the given pipeline	The logs page of a pipeline	Error shown	Added Priority, Failure

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_102	T3.4 & T3.5	M	Data Scientist, Data Engineer	view detailed execution logs for an AI pipeline	I can detect bottlenecks once my AI pipeline has been executed	The user can see the logs output for a given pipeline run	The execution logs for a given pipeline run	The operation fails with a visible error on the page	Added Priority
TR_103	T3.4 & T3.5	M	Data Scientist, Data Engineer	keep a history of experiments performed (pipeline runs)	I can compare results & improve the pipeline	The user can see all runs for a pipeline	A list of runs for a pipeline	Error shown	Added Priority, Failure
TR_104	T3.4 & T3.5	H	Data engineer	schedule the execution of the AI pipelines	I can have at my disposal up-to-date results	The user can schedule a pipeline for execution	The pipeline schedules successfully as shown in the logs page	The scheduling fails with errors visible in the logs page	Added Priority
TR_105	T3.4 & T3.5	M	Business User, Data Scientist	store the execution results	I can retrieve them to use them in external systems	The user can export execution results in standardized format	A standardized file format containing the execution results	Error shown	Added Priority, Failure
TR_106	T3.4 & T3.5	Deprecated	Data Engineer	receive notifications when certain metrics exceed defined thresholds	I can timely investigate problems in deployed pipeline	-	-	-	Deprecated requirement
TR_107	T3.4	M	Data Scientist, Data Engineer	have the required resources and automatic parallelization, when dealing with big data manipulation	I can execute an AI pipeline in a faster and more reliable way	The pipeline performs parallel executions, where possible	A parallelized pipeline, when applicable	Error shown	Added Priority, Failure

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_108	T3.5	M	Business user, Data Scientist	execute AI pipelines locally on my private cloud or servers on premise	I can perform my analysis in a secure and trusted environment	If a user wants to execute a designed AI pipeline on his/her on premise environment, the pipeline is started from a local deployment of the execution engine	the user can select and execute an AI pipeline locally on his/her on premise environment	The user cannot execute an AI pipeline locally on his/her on premise environment	Added Priority
TR_109	T3.5	M	Data Scientist	create reports of the performed data analysis locally on my environment	I can evaluate the results	If a user wants to create a visualisation of an existing result on his/her on premise environment, the user selects it and generates it locally.	the user can select a locally available result and visualise it	the user cannot visualise the locally available results	Added Priority
TR_110	T3.4 & T3.5	H	Data Scientist	export my analysis results or AI models as files	I can import them on my organisation's systems	the user is able to select and export the results that he/she is eligible to export as a file	the user selects a result from the list of eligible for exporting results and exports it as a file	the user cannot export as a file a result from the list of eligible for exporting results	Added Priority
TR_111	T3.4 & T3.5	H	Data Scientist	retrieve my analysis results through an API	I can use them in the XMANAI manufacturing apps	the user can export the results that he/she is eligible to export via an API	the user initiates an API call and retrieves the result that he/she is eligible to retrieve	The user cannot retrieve the results that he/she is eligible to retrieve via API	Added Priority
TR_112	T3.4 & T3.5	M	Data Engineer	containerize the developed AI pipelines	they can be more easily deployed	The pipeline gets executed in a containerized environment	A pipeline running in a containerized environment	Error shown	Added Priority, Failure
TR_113	T3.5	H	Data Scientist	able to setup a data analysis execution and results visualization environment easily on private cloud or servers on premise	I can leverage the execution of the analysis and visualization of results on infrastructures I control	The on-premise environment is easily installed on a private cloud or server	The user is able to easily setup the on-premise environment on their own private cloud or server	The on-premise environment cannot be installed on a private cloud or server	Added Priority

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_114	T4.1 & T4.2	Deprecated	Data Scientist	Establish a common naming protocol for data files uploaded to the XMANAI platform	Specific information on the file contents can be deduced from the filename.	Basic facts regarding the object under analysis and the scanning setup are exposed in the filename.	It is easy to understand the contents and manipulate measurement files based on the measured object's geometry and the scanning parameters.	The filename does not point to the correct geometry, dimension and scanning parameters.	Deprecated as it's not a generic requirement, but a (single) demonstrator-specific one
TR_115	T4.1 & T4.2	Deprecated	Data Scientist	Define a file/data format protocol to transform information contained in CAD files of the objects to be analysed	Information on the objects' nominal design can be ingested from the M3 metrological software to the XMANAI platform and become available for analysis & experimentation.	A nominal Point Cloud for the object under study is ingested in the XMANAI platform.	A prediction can be performed based on the nominal Point Cloud of the object under study.	An initial scan with randomly selected scanning parameters must be performed, to obtain an initial Point Cloud and perform the prediction.	Deprecated as it's not a generic requirement, but a (single) demonstrator-specific one
TR_116	T4.1 & T4.2	Deprecated	Data Scientist	Define a file/data format protocol to transform GD&T information contained in QIF files of the objects to be analysed	Information on GD&T regarding the measurement can be ingested from M3 metrological to XMANAI platform and be available for analysis & experimentation.	The target of the measurement (dimensional/positional/both) and associated tolerances are accessible in the platform.	Data scientist/ Data Engineer can access the GD&T information through the platform to apply the appropriate models depending on the target of the measurement.	The target & tolerances of the measurement are not defined.	Deprecated as it's not a generic requirement, but a (single) demonstrator-specific one

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_117	T4.1 & T4.2	M	Data Scientist	provide wrappers (register) for explainability techniques	I can include them in my AI pipeline	Available common interfaces to interact with explainability tools	To be able to execute these explainability tools through these wrappers	Error shown	Added Priority and Failure. The wrappers refer to specific configuration that needs to be in place.
TR_118	T4.1 & T4.2 (& T3.3, T2.3)	M	Data Scientist	have a common metadata model for describing my trained models (hyperparameters, parameters, code, metrics...)	I can share them with other users	Creation of a common metadata model for ML models	Common metadata model for ML models	Error shown	Added Priority and Failure.
TR_119	T4.1 & T4.2 (& T3.3, T2.3)	M	Data Scientist	package and make available the trained AI models I want to execute (following specific guidelines for directory tree, programming language, name of files, etc.)	I can upload them to the XMANAI catalogue	Availability to package a model to the platform	Model available in the platform	Error shown	Added Prioroty. Updated Description, Success to clarify better the scope.
TR_120	T4.1, T4.2 & T3.3	M	Data Scientist	register a trained AI model I have created offline to solve a specific problem	I can make it available to be reused by other users	Given a user wants to register an offline trained AI model, when uploaded to the Assets Store, then I can register it in a session to be used to perform inferences	Model registered into the platform	Error shown	Added Priority

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_121	T4.1, T4.2 & T3.3, T2.4	M	Data Scientist	keep versioning for the trained AI model I have created offline	I can retrain the model without affecting all AI pipelines it has been reused	Given a user wants to retrain an offline trained model uploaded to the platform, when the model is uploaded and the pipelines are run, then the user will get a new trained model, not overwriting the previous one, without affecting all previous pipelines	A new model will be created that will not overwrite the previous one	Error Shown	Added Priority
TR_122	T4.1, T4.2 & T3.3	M	Data scientist	follow specific guidelines for the explainability techniques (e.g. surrogate models) I want to register	all techniques and their associated metadata (including python packages requirements) can be packaged and available in the AI pipelines	To provide guidelines to accomplish with a set of requirements	Defined guidelines to follow when an explainability tool has to be registered	Error Shown	Added Priority
TR_123	T4.1, T4.2 & T3.3	M	Data Scientist	register a trained explainability technique I have created manually	I can use it in my AI pipelines	Given the user wants to register a new explainability technique created offline, when he has created it, then he can upload it through a frontend available in the XMEE component	Model registered into the platform	Error shown	Added Priority. Adapted Description.
TR_124	T4.3	M	Data Scientist	know which validation method is right for my AI model	I can validate the model(s) with higher confidence	type of data and number of observations	comply with the rule as have been decided in the validation protocol	Not enough data (observations) and/or features	Added Priority

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_125	T4.3 (T4.1, T4.2, T3.3)	H	Data Scientist	use various performance and evaluation metrics for the AI models that satisfy both data scientists and business users	I have a better picture of my models' effectiveness	number of performance metrics per AI model	more than 3 performance metrics	less than 3 performance metrics	Added Priority
TR_126	T4.3 (T3.3)	M	Data Scientist	generate an evaluation report for each trained ML/DL model	users that insert it in their AI pipelines are aware of its performance	evaluation report per trained model	existence	absence	Added Priority
TR_127	T4.3	H	Business User (Expert)	evaluate the validity of the explanations and provide feedback	I can improve the AI models and AI pipelines to solve a specific problem	number of XAI tools used	at least 2 XAI tools	none	Added Priority
TR_128	T4.3 (T3.3)	M	Data Scientist	compare results from different models trained on different features for a particular task	I can gain an understanding of which factors aid in the performance of the models (which features helped, how the different pre-processing steps affected the result	number of models	at least three	less than three	Added Priority
TR_129	T4.3 (T3.3)	L	Data Scientist	compare performances of different AI pipelines	I can track the improvements of the different versions of my pipelines / models	performance	improvement over previous versions	No improvement comparing to previous versions	Added Priority

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_130	T4.3	H	Business User	view reports of the experiments (simulations of different settings, models, and methods) to solve a specific problem	I can evaluate the results	evaluation report per trained model	existence	absence	Added Priority
TR_131	T4.4	M	Data Scientist	identify any possible corruption of the stored model	I can check the integrity of the stored model	The model owner can perform integrity checks on the stored models	Integrity checks are performed and possible alternations are identified	No integrity checks can be performed or alternations cannot be identified.	Added Priority
TR_132	T4.4	M	Data Scientist	check the stored model against possible adversarial attacks	I can repair possible poisoning of the model	The user can check if adversarial attack has been performed on the model	Checks against adversarial attacks are performed and possible poisoning is identified	Checks against adversarial attacks cannot be performed or possible adversarial attacks cannot be identified	Added Priority
TR_133	T4.4	Deprecated	Data Scientist, Data Engineer	generate adversarial examples	I can create a robust model against adversarial attacks	-	-	-	Deprecated requirement
TR_134	T4.4	Deprecated	Data Scientist	check training data sets for possible unfair biases	I can prevent possible discriminatory biases of the model	-	-	-	Deprecated requirement

Nº	Task	Priority	As a…	I want to…	So that…	Acceptance criteria	Success	Failure	Update
TR_135	T2.2	H	Business user, Data Scientist	able to define the access level of my assets only to selected organisations with specific attributes	I can define which organisations shall have access to my assets and under which circumstances	The asset owner is a capable of selecting the attributes of the organisations (from a predefined list of attributes) that can access each asset	Specific attributes of the organisation that can access the asset is properly defined	The asset owner cannot define the attributes of the organisations that shall have access to each asset.	Added Priority

Following the ISO/IEC 25010:2011, the "Non-functional" requirements define system attributes such as security, reliability, performance, maintainability, scalability and usability. Also known as system qualities, they are just as critical as functional requirement as they safeguard the usability and effectiveness of the entire system. Table 3-2 presents these requirements that failing to meet any of them can result in systems, that fail to satisfy business or markets or user needs.

Table 3-2: Final XMANAI Non-functional requirements

ID	Description	Category
NF_Req_01	The XMANAI platform should have high availability	Reliability
NF_Req_02	XMANAI should support know-your-customer practices, with organization registration and user login with credentials	Security
NF_Req_03	XMANAI should be able to execute XAI pipelines in a timely and efficient manner	Performance efficiency
NF_Req_04	XMANAI should guarantee the efficient and effective resource allocation for the success analytics jobs execution	Performance efficiency
NF_Req_05	XMANAI should be able to support the functional and flexible operation in a distributed cloud infrastructure	Compatibility
NF_Req_06	XMANAI should be able to consume and handle different datasets in various formats	Compatibility
NF_Req_07	XMANAI should provide an easy-to-use and user-friendly interface in which the analytics and visualisation processes are supported by guides and manuals	Usability
NF_Req_08	XMANAI should provide the suitable error protection methods for all input fields	Usability
NF_Req_09	XMANAI should enable the secure storage of assets (datasets, reports, etc.)	Reliability
NF_Req_10	XMANAI should be able to handle simultaneous requests on a timely and efficient manner	Reliability

ID	Description	Category
NF_Req_11	XMANAI should provide the mechanisms to recover after system failure conditions	Reliability
NF_Req_12	XMANAI should be able to handle software errors without affecting the platform overall functionality	Reliability
NF_Req_13	XMANAI should ensure different authorisation access to different datasets	Security
NF_Req_14	XMANAI should provide the appropriate logging mechanisms for all operations	Security
NF_Req_15	XMANAI should be able to verify the identity of the user/subject performing any operation	Security
NF_Req_16	XMANAI should be composed by components that are operating independently	Maintainability
NF_Req_17	XMANAI should provide the proper mechanisms for system upgrade with minimum downtime	Maintainability
NF_Req_18	XMANAI should offer easy installation process for the On-Premise Environments in a timely manner	Portability
NF_Req_19	XMANAI should be composed by independent components that are replaceable with minimum impact and effort	Portability

3.3 Technical Requirements vs Business Requirements

A correlation between the technical requirements and the business requirements (extracted from D6.3) to highlight how the business requirements have been considered. It needs to be noted that there might be multiple technical requirements relevant for each business requirement. This table is updated from D1.3 according to the updates of the technical requirements.

Table 3-3: Business – Technical Requirements Alignment

Business Requirements		Technical Requirements No.
No.	**Description**	
BR_1	When a failure affect occurs, XMANAI shall represent in real time or near real time the machine/part related to the failure.	TR_65, TR_66, TR_67, TR_68, TR_70, TR_97, TR_114, TR_115, TR_116
BR_2	XMANAI shall send advice based on real time data to help to take actions about the stocks.	TR_17, TR_18, TR_19, TR_65, TR_66, TR_67, TR_68
BR_3	XMANAI should integrate all corporative data to know the status of machines.	TR_1, TR_3, TR_16, TR_42
BR_4	XMANAI should allow operators and engineers to visualize the data from any locations and from multiple devices (smartphone, laptops).	TR_20, TR_51, TR_52, TR_53, TR_96, TR_98, TR_108, TR_113
BR_5	XMANAI should save historical data to help engineers to review the historical actions.	TR_45, TR_103
BR_6	XMANAI shall show direct and visual alarms to alert about critical situations	TR_73, TR_76, TR_97, TR_101, TR_106
BR_7	XMANAI should allow an operator to understand the root causes in every working situation.	TR_75, TR_95, TR_97, TR_127
BR_8	XMANAI should have flexibility to be applied in different lines.	TR_65, TR_66, TR_67, TR_68, TR_93, TR_118, TR_123
BR_9	XMANAI should provide advice in terms of production plan.	TR_65, TR_66, TR_67, TR_68, TR_84, TR_89, TR_91, TR_92, TR_117, TR_127
BR_10	XMANAI should read and integrate data from corporate databases and external sources.	TR_1, TR_3, TR_16, TR_32, TR_42, TR_43
BR_11	XMANAI should act as simulator of different planning scenarios.	TR_65, TR_66, TR_67, TR_68, TR_123, TR_128, TR_129, TR_130
BR_12	XMANAI should consider the production that is currently in the line but it hasn't yet finished.	TR_1, TR_2, TR_16
BR_13	XMANAI should be agile to replan when an unexpected event occurs in the line.	TR_65, TR_66, TR_67, TR_68, TR_77, TR_97, TR_100, TR_106,
BR_14	XMANAI should provide measurements of the deviation between the predicted plan and the real production.	TR_87, TR_94, TR_124, TR_125, TR_126, TR_127
BR_15	XMANAI should alert for critical parts, that sufficient stocks of some parts to finish the plan are not available.	TR_97, TR_101, TR_106
BR_16	XMANAI should allow a central planner to have a correct forecasting of D2C sales per day/product in a horizon of 3 months.	TR_65, TR_66, TR_67, TR_68, TR_70, TR_85, TR_114, TR_115, TR_116, TR_120, TR_130
BR_17	XMANAI should generate demand forecasts on a daily basis.	TR_65, TR_66, TR_67, TR_68, TR_70, TR_85, TR_114, TR_115, TR_116, TR_120, TR_130

Business Requirements		Technical Requirements No.
No.	Description	
BR_18	XMANAI shall allow central planners/D2C marketing, D2C sales/D2C logistics to visualise the key factors effects influencing demand profile.	TR_65, TR_66, TR_67, TR_68, TR_88, TR_89, TR_90, TR_114, TR_115, TR_116, TR_117
BR_19	XMANAI shall allow central planners/D2C marketing, D2C sales/D2C logistics to see the clustering of customer behavior.	TR_66, TR_88, TR_89, TR_90, TR_115, TR_116
BR_20	XMANAI shall allow central planners/D2C marketing, D2C sales/D2C logistics to visualise buying patterns per customer profile/product/period.	TR_65, TR_66, TR_67, TR_68, TR_78, TR_84, TR_88, TR_89, TR_90, TR_114, TR_115, TR_116, TR_117
BR_21	XMANAI shall allow D2C marketing/D2C sales to receive recommendations/input for promotional actions.	TR_65, TR_66, TR_67, TR_68, TR_78, TR_84, TR_114, TR_115, TR_116
BR_22	XMANAI shall allow central planners/D2C marketing, D2C sales/D2C logistics to simulate demand forecasting forcing the change in one or more key parameters.	TR_65, TR_66, TR_67, TR_68, TR_70, TR_75, TR_79, TR_85, TR_114, TR_115, TR_116, TR_130
BR_23	XMANAI shall strictly authorize users for system access.	TR_21, TR_22, TR_23, TR_24, TR_25, TR_27, TR_28, TR_31, TR_45, TR_46, TR_57, TR_58
BR_24	XMANAI shall strictly protect all sales data through encryption and secure data management, in compliance with Whirlpool data security policies.	TR_20, TR_21, TR_22, TR_23, TR_24, TR_25 , TR_27, TR_28, TR_31, TR_46, TR_57, TR_58, TR_108
BR_25	XMANAI shall provide full customer data anonymization	TR_29, TR_30, TR_57, TR_58
BR_26	XMANAI shall fully respect GDPR (General Data Protection Regulation) by a privacy by design approach, in compliance with Whirlpool data security policies.	TR_21, TR_22, TR_23, TR_24, TR_25, TR_27, TR_28, TR_29, TR_30, TR_31, TR_45, TR_57, TR_58,
BR_27	XMANAI should provide an interactable digital twin able to forecast the behaviour of the machinery.	TR_65, TR_66, TR_67, TR_68, TR_69, TR_70, TR_114, TR_115, TR_116, TR_117, TR_120, TR_130
BR_28	XMANAI should provide an interactable digital twin able to improve the collaboration and communication between production parts and engineering CAD model of the parts themselves.	TR_64, TR_69, TR_80, TR_83, TR_91, TR_92
BR_29	XMANAI should provide suggestions with description of the problem and visibility on how the problem has been identified.	TR_83, TR_84, TR_88, TR_89, TR_90, TR_97, TR_101, TR_109
BR_30	XMANAI should provide an interactable HMI able to improve the comprehension of the suggestion provided by the XAI and to navigate them.	TR_64, TR_77, TR_78, TR_84, TR_89, TR_91, TR_92, TR_93, TR_118, TR_127
BR_31	XMANAI should provide a set of data to the user identifying the problem before the critical moment, the maintenance/troubleshooting procedure to be executed and the parameters to be monitored during production.	TR_17, TR_18, TR_19, TR_75, TR_97, TR_106, TR_110, TR_111
BR_32	XMANAI should support the Blue Collar Worker is doing the maintenance/troubleshooting procedures with AR/XAI connection.	TR_65, TR_66, TR_67, TR_68, TR_70, TR_77, TR_78, TR_89, TR_91, TR_92, TR_114, TR_115, TR_116, TR_117, TR_127
BR_33	XMANAI should learn from PMS past quality problems and parameters which was leading to the problem.	TR_65, TR_66, TR_67, TR_68, TR_70, TR_97, TR_114, TR_115, TR_116
BR_34	XMANAI should provide product quality risks starting from production parameters including explanation of why certain conditions happened or are forecasted to happen.	TR_77, TR_82, TR_88, TR_89, TR_90, TR_91, TR_92 TR_109, TR_125, TR_126, TR_127, TR_130

Business Requirements		Technical Requirements No.
No.	Description	
BR_35	XMANAI should generate a report with the quality risks for the production manager.	TR_82, TR_88, TR_89, TR_90, TR_109, TR_125, TR_126, TR_127, TR_130
BR_36	XMANAI shall be able to adjust the displayed criteria based on the geometry type and allows full control to add various GD&T checks and other specific location information for an element.	TR_53, TR_75, TR_79, TR_85, TR_94
BR_37	XMANAI shall be able to automatically add and connect to an instrument using predefined parameters without any user interaction after configuration starts.	TR_70, TR_86, TR_104, TR_112
BR_38	XMANAI shall reduce the amount of interaction with the software so that users spend more time measuring and less time browsing through the software.	TR_70, TR_71, TR_86, TR_100, TR_104, TR_107, TR_112, TR_113, TR_120, TR_122, TR_123
BR_39	XMANAI should collect and standardize machining data.	TR_1, TR_3, TR_10, TR_11, TR_16, TR_12, TR_13, TR_93, TR_118,
BR_40	XMANAI shall be able to provide information to detect easily the sources of problems.	TR_75, TR_89, TR_90, TR_97, TR_101, TR_102, TR_106, TR_110, TR_111
BR_41	XMANAI shall create pop-up messages for user instructions, instrument alignment, profile change, etc.	TR_76, TR_97, TR_101, TR_106
BR_42	XMANAI shall keep historical records of the machine on-site.	TR_45, TR_102, TR_103
BR_43	XMANAI shall be able to analyze and communicate results.	TR_63, TR_69, TR_80, TR_91, TR_92, TR_95, TR_126, TR_127, TR_130
BR_44	XMANAI shall create predefined measurement routines and explain why they are the correct ones.	TR_5, TR_6, TR_7, TR_8, TR_9, TR_94
BR_45	XMANAI shall be able to reduce the number of iterations required with the computer by the user, so the data collection will be more efficient.	TR_2, TR_16, TR_54, TR_64
BR_46	XMANAI shall perform improved visualizations of item annotations and value logic.	TR_83, TR_84, TR_88, TR_89, TR_90

3.4 Technical Requirements across the User Journeys

Table 3-4 shows the alignment of technical requirements across the Data Scientist User Journey. This table has been updated based on the changes in the technical requirements from D1.3.

Table 3-4: Technical requirements alignment across Data Scientist User Journey

User Journey Step	Technical Requirement No.
Understand the data and the problem at hand	TR_1, TR_2, TR_3, TR_5, TR_6, TR_7, TR_8, TR_9, TR_10, TR_11, TR_12, TR_13, TR_14, TR_15, TR_16, TR_17, TR_18, TR_19, TR_21, TR_22, TR_23, TR_24, TR_25, TR_27, TR_28, TR_29, TR_30, TR_31, TR_32, TR_35, TR_36, TR_37, TR_39, TR_40, TR_42, TR_43, TR_44, TR_50, TR_51, TR_52, TR_53, TR_54, TR_107
Prepare the data and handle problematic data cases	TR_41, TR_55, TR_56, TR_57, TR_58, TR_59, TR_60, TR_61, TR_62, TR_131, TR_132,
Collaboratively design XAI pipelines	TR_65, TR_66, TR_67, TR_68, TR_69, TR_70, TR_71, TR_73, TR_75, TR_76, TR_77, TR_78, TR_79, TR_80, TR_81, TR_83, TR_86, TR_96, TR_101, TR_102, TR_114, TR_115, TR_116, TR_117

Explain XAI models/ results	TR_82, TR_85, TR_87, TR_88, TR_92, TR_93, TR_103, TR_105, TR_108, TR_109, TR_110, TR_111, TR_113, TR_118, TR_119, TR_120, TR_121, TR_122, TR_123
Evaluate XAI models/ pipelines	TR_97, TR_98, TR_99, TR_100, TR_124, TR_125, TR_126, TR_128, TR_129

Table 3-5 shows the alignment of technical requirements across the Data Engineer User Journey. This table has been updated based on the changes in the technical requirements from D1.3.

Table 3-5: Technical requirements alignment across Data Engineer User Journey

User Journey Step	Technical Requirement No.
Collaborate on uploading data/Handling data	TR_1, TR_2, TR_3, TR_4, TR_5, TR_6, TR_7, TR_8, TR_9, TR_10, TR_11, TR_12, TR_13, TR_14, TR_15, TR_16, TR_19, TR_20, TR_28, TR_29, TR_30, TR_32, TR_34, TR_35, TR_36, TR_37, TR_39, TR_40, TR_42, TR_44, TR_107
Collaborate on the design of XAI pipelines	TR_65, TR_66, TR_67, TR_69, TR_70, TR_96, TR_101, TR_102
Deploy XAI pipelines to production	TR_70, TR_71, TR_73, TR_82, TR_103, TR_104, TR_106, TR_108, TR_112

Table 3-6 shows the alignment of technical requirements across the Business User/User Journey. This table has been updated based on the changes in the technical requirements from D1.3.

Table 3-6: Technical requirements alignment across Business User/User Journey

User Journey Step	Technical Requirement No.
Provide data	TR_1, TR_2, TR_3, TR_12, TR_13, TR_14, TR_15, TR_16, TR_20, TR_21, TR_22, TR_23, TR_24, TR_25, TR_27, TR_28, TR_29, TR_30, TR_31, TR_32, TR_34, TR_36, TR_37, TR_39, TR_40, TR_45, TR_46, TR_47, TR_48, TR_49
Contribute to better understanding and evaluating data/ XAI models / Results	TR_35, TR_63, TR_64, TR_127
Gain insights into XAI results	TR_84, TR_89, TR_90, TR_91, TR_94, TR_95, TR_105, TR_130

…

4 XMANAI Final Minimum Viable Product (MVP)

This section describes the approach followed step-by-step in order to finalize the XMANAI Minimum Viable Product (MVP) taking into consideration the project advancements since the early MVP release that was documented in the XMANAI Deliverable D1.2 and the intermediate MVP release described in the XMANAI Deliverable D1.3.

4.1 Overview

As already discussed in D1.2, the Minimum Viable Product (MVP) refers to a version of a product with the minimum set of features and functionalities that can satisfy early adopters who, in turn, can promptly provide feedback for future product improvements. For XMANAI, the MVP represents the overall mindset and strategy adopted for distributing efficiently the development and integration workload, for continuously testing the end user reaction, and validating the methodological ideas and hypothesis. This can easily be depicted in the following figure, Figure 4-1, which presents in brief the XMANAI MVP definition approach, involves three core phases, namely Feature Definition, Feature Assessment and MVP Consolidation and runs over the three iterations of WP1.

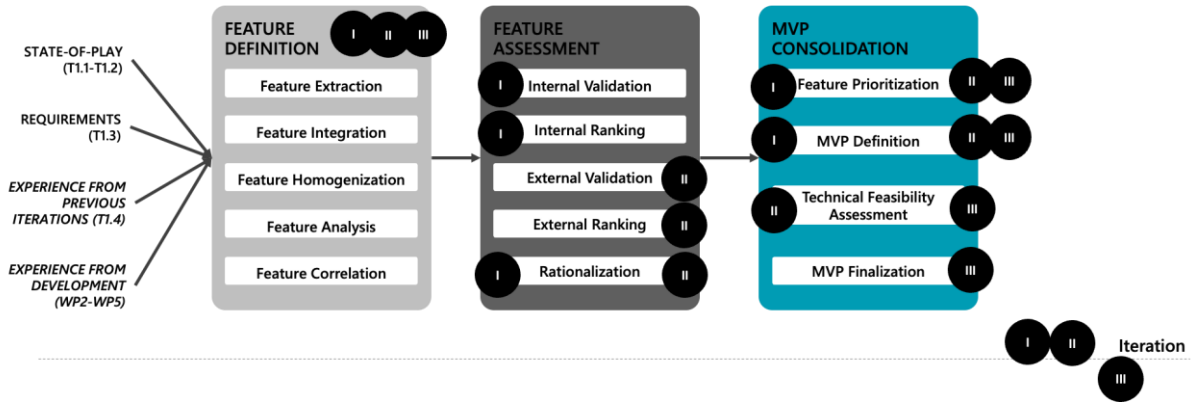

Figure 4-1: XMANAI MVP Approach

During the final iteration, the XMANAI MVP builds on its previous releases and incorporates the experience gained from the development activities (in WP2, WP3, WP4 and WP5) in its implicit technical feasibility assessment while the feedback acquired from the early demonstration activities (in WP6) has also provided additional insights considered for the final feature prioritization.

Following the MoSCoW approach (that classifies the requirements into “Must-have”, “Should-have”, “Could-have”, and “Won't-have”, or “Will not have right now”), the set of features that comprised the preliminary and intermediate XMANAI MVP were defined. However, for the final MVP release, the features were re-evaluated to reflect the project’s advancements and instead of setting their planned priority (as in the previous MVP releases), their actual status (in terms of “Will-Have” and “Won’t-Have”) is reflected.

It needs to be noted that the detailed results for the feature assessment have been provided in the XMANAI Deliverable D1.2 and D1.3 and are not repeated in this document.

4.2 MVP Features In-Depth Analysis

During the final iteration, the set of MVP features has been refined and updated taking into consideration their definition in the early and intermediate MVP releases and the final technical requirements (described in Section 3), in order to appropriately reflect the current project status, orientation and planning. The features of the final MVP are classified under the following categories: Data Integration (DI), Data Management (DM), Data Sharing (DS), Data Preparation (DP), Artificial Intelligence (AI), Explainability (EX), and User Management (UM), as presented below.

XMANAI_F_UM_001. Organization-based access

Description	In the XMANAI platform, users always belong to one organization (group) and are invited by their organization's manager. They can view and use all data assets that belong to their organization or to which their organization has access. Users within the same organization are also able to collaborate by accessing the pipelines that other users within their organization have created.
Category	I. User Management
Prerequisites	-
Status	Will-Have
Changes Since Previous Releases	Updated the feature description. Updated the Priority "Must-have" (which the feature had in the intermediate MVP release) to Status "Will-Have"

XMANAI_F_UM_002. Project-based access

Description	In XMANAI platform, users from different organizations can access data assets that are grouped under a project in order to ensure that users from different organizations can contribute for a common purpose/business problem.
Category	I. User Management
Prerequisites	-
Status	Won't Have
Changes Since Previous Releases	No change since it was confirmed that the specific feature will not be supported

XMANAI_F_UM_003. Delete all data assets

Description	The XMANAI platform should allow for the deletion of all data assets and eventually the account that belongs to an entity (either at organization level or at individual/user level) as per GDPR guidelines regarding the 'right to be forgotten'.
Category	I. User Management
Prerequisites	-
Status	Will-Have
Changes Since Previous Releases	Updated the feature description. Updated the Priority "Could-have" (which the feature had in the intermediate MVP release) to Status "Will-Have"

XMANAI_F_DI_004. Data Sources Management

Description	The XMANAI platform should allow users to manage various types of data sources (files, APIs, etc) out of which data will be ingested. Each data source will be handled independently, allowing for its custom configuration, creation, update, deletion and scheduling its execution, as needed.
Category	II. Data Ingestion (User Journey(s): Business User Phase 1, Data Scientist Phase 1, Data Engineer Phase 1)
Prerequisites	• A connector/importer to data source is already available in XMANAI • The XMANAI Data model is sufficient to describe the data from the data source
Status	Will-Have

Changes Since Previous Releases	Updated the feature description. Updated the Priority "Must-have" (which the feature had in the intermediate MVP release) to Status "Will-Have"

XMANAI_F_DI_005. Data Secure Uploading as file(s)

Description	The XMANAI platform should allow for the safe and reliable uploading of batch data as a file that follows specific formats (e.g. csv).
Category	II. Data Ingestion (User Journey(s): Business User Phase 1, Data Scientist Phase 1, Data Engineer Phase 1)
Prerequisites	-
Status	Will-Have
Changes Since Previous Releases	Updated the feature description. Updated the Priority "Must-have" (which the feature had in the intermediate MVP release) to Status "Will-Have"

XMANAI_F_DI_006. Data Secure Uploading via API

Description	The XMANAI platform should allow for the safe and reliable uploading of real-time data via an API that is already exposed by an organization.
Category	II. Data Ingestion (User Journey(s): Data Scientist Phase 1, Data Engineer Phase 1)
Prerequisites	• A connector/importer to the specific API is already available in XMANAI
Status	Will-Have
Changes Since Previous Releases	Updated the feature description. Updated the Priority "Must-have" (which the feature had in the intermediate MVP release) to Status "Will-Have"

XMANAI_F_DI_007. Data Mapping to a Data Model and Harmonization

Description	The XMANAI platform should provide a common data model to which all uploaded datasets should adhere, in order to ensure explainability at data level. This mapping process will ensure commonly agreed semantics and perform harmonization techniques, i.e. data type casting.
Category	II. Data Ingestion (User Journey(s): Business User Phase 1, Data Scientist Phase 1, Data Engineer Phase 1)
Prerequisites	• Appropriate categories for the data have been foreseen in the data model.
Status	Will-Have
Changes Since Previous Releases	Updated the feature description and the related XMANAI Components. Updated the Priority "Should-have" (which the feature had in the intermediate MVP release) to Status "Will-Have"

XMANAI_F_DI_008. Data Model Management

Description	The XMANAI platform should allow for managing a data model by an administrator in order to add/edit/remove its contents (i.e. concepts, properties, relationships), and handle any suggestions provided. All users should be able to navigate to the data model's categories, view its concepts, properties, and relationships, and suggest any additions that are relevant for their data.
Category	II. Data Ingestion (User Journey(s): Business User Phase 1, Data Scientist Phase 1)

Prerequisites	-
Status	Will-Have
Changes Since Previous Releases	Updated the feature description and the related XMANAI Components. Updated the Priority "Must-have" (which the feature had in the intermediate MVP release) to Status "Will-Have"

XMANAI_F_DI_009. Data Cleansing

Description	The XMANAI platform should provide a cleansing mechanism and allow for quality checks and the definition and execution of cleaning rules before storing the data.
Category	II. Data Ingestion (User Journey(s): Business User Phase 1, Data Scientist Phase 1, Data Engineer Phase 1)
Prerequisites	-
Status	Won't Have
Changes Since Previous Releases	No change since it was confirmed that the specific feature will not be supported before the data are stored.

XMANAI_F_DI_010. Data Anonymisation

Description	The XMANAI platform should provide an anonymization mechanism and allow for the definition and execution of anonymization rules before uploading/storing the data. The anonymization functionality is stand-alone, offered on-premise.
Category	II. Data Ingestion (User Journey(s): Business User Phase 1, Data Scientist Phase 1, Data Engineer Phase 1)
Prerequisites	-
Status	Will-Have
Changes Since Previous Releases	Updated the Priority "Could-have" (which the feature had in the intermediate MVP release) to Status "Will-Have"

XMANAI_F_DI_011. Data Storage in central XMANAI Cloud Storage

Description	The XMANAI platform should allow for the secure storage of data assets (including datasets, models, results, files) in centralized servers in the cloud (in the XMANAI Centralized Cloud layer).
Category	II. Data Ingestion (User Journey(s): Business User Phase 1, Data Scientist Phase 1, Data Engineer Phase 1)
Prerequisites	-
Status	Will-Have
Changes Since Previous Releases	Updated the feature description. Updated the Priority "Must-have" (which the feature had in the intermediate MVP release) to Status "Will-Have"

XMANAI_F_DI_012. Data Storage on-premise / in private cloud

Description	The XMANAI platform should allow for the secure storage of data assets (including datasets, models, results, files) in on-premise private cloud spaces that are available per demonstrator (in the XMANAI Stakeholders' On-Premise layer).

Category	II. Data Ingestion (User Journey(s): Business User Phase 1, Data Scientist Phase 1, Data Engineer Phase 1)
Prerequisites	Appropriate infrastructures (in terms od processing and memory) need
Status	Will-Have
Changes Since Previous Releases	Updated the feature description. Updated the Priority "Must-have" (which the feature had in the intermediate MVP release) to Status "Will-Have"

XMANAI_F_DI_013. Data Encryption

Description	The XMANAI platform should allow for the encryption of data while a) transferred onto the platform and/or b) stored on the platform
Category	II. Data Ingestion (User Journey(s): Business User Phase 1, Data Scientist Phase 1, Data Engineer Phase 1)
Prerequisites	-
Status	Won't Have
Changes Since Previous Releases	No change since it was confirmed that the specific feature will not be supported.

XMANAI_F_DM_014. Dataset Management

Description	The XMANAI platform should provide the means for managing data assets, including adding a new dataset, including its metadata, updating its data, or removing it. Keeping different versions of a dataset for provenance purposes is also part of this feature.
Category	III. Data Asset Management and Security (User Journey(s): Business User Phase 2, Data Scientist Phase 2, Data Engineer Phase 1)
Prerequisites	XMANAI_F_DI_011
Status	Will-Have
Changes Since Previous Releases	Updated the feature description. Updated the Priority "Must-have" (which the feature had in the intermediate MVP release) to Status "Will-Have"

XMANAI_F_DM_015. Features Management

Description	The XMANAI platform should provide functionalities for storing and managing curated features within a feature store, allowing for the addition, editing or removal of features, appending metadata to better describe these assets and keeping versions of features.
Category	III. Data Asset Management and Security (User Journey(s): Data Scientist Phase 2, Data Engineer Phase 1)
Prerequisites	-
Status	Won't Have
Changes Since Previous Releases	No change since it was confirmed that the specific feature will not be supported.

XMANAI_F_DM_016. Results Management

Description	The XMANAI platform should support the creation and storage of results generated by a XAI pipeline. Appending metadata to better describe these assets and potentially keeping versions of them can also be part of this feature.
Category	III. Data Asset Management and Security (User Journey(s): Data Scientist Phase 2, Data Engineer Phase 1)
Prerequisites	-
Status	Will-Have
Changes Since Previous Releases	Updated the feature description and the related XMANAI components. Updated the Priority "Should-have" (which the feature had in the intermediate MVP release) to Status "Will-Have"

XMANAI_F_DM_017. AI Model Management

Description	The XMANAI platform should provide model management functionalities to configure, create, train, store, update, delete AI models, while keeping separate versions of each configuration and experiment associated with them.
Category	III. Data Asset Management and Security (User Journey(s): Data Scientist Phase 3, Data Engineer Phase 2)
Prerequisites	-
Status	Will-Have
Changes Since Previous Releases	Updated the feature description. Updated the Priority "Must-have" (which the feature had in the intermediate MVP release) to Status "Will-Have"

XMANAI_F_DM_018. AI Pipeline Management

Description	The XMANAI platform should allow for managing XAI Pipelines including their creation, configuration, storage, update, and deletion. In addition, the platform should allow the inclusion of different assets (datasets, models in terms of training sessions, results) in an XAI pipeline taking into consideration their IPRs.
Category	III. Data Asset Management and Security (User Journey(s): Data Scientist Phase 3, Data Engineer Phase 2)
Prerequisites	-
Status	Will-Have
Changes Since Previous Releases	Updated the feature description and the Related XMANAI Components. Updated the Priority "Must-have" (which the feature had in the intermediate MVP release) to Status "Will-Have"

XMANAI_F_DM_019. Data Asset Export

Description	The XMANAI platform should allow for exporting datasets and results in different ways (e.g. via file download or via APIs), depending on the IPR.
Category	III. Data Asset Management and Security (User Journey(s): Data Scientist Phase 3, Data Engineer Phase 2)
Prerequisites	XMANAI_F_DI_011
Status	Will-Have

Changes Since Previous Releases	Updated the feature description and the Related XMANAI Components. Updated the Priority "Should-have" (which the feature had in the intermediate MVP release) to Status "Will-Have"

XMANAI_F_DM_020. Data Asset Access Policies and Licensing

Description	The XMANAI platform should allow users to manage and configure the access policies and licensing rules that they want to enforce for their data assets (datasets, models, results) at run-time across all their operations.
Category	III. Data Asset Management and Security (User Journey(s): Business User Phase 2, Data Scientist Phase 3)
Prerequisites	-
Status	**Will-Have**
Changes Since Previous Releases	Updated the feature description and the Related XMANAI Components. Updated the Priority "Must-have" (which the feature had in the intermediate MVP release) to Status "Will-Have"

XMANAI_F_DM_021. Data Asset Secure Transfer

Description	The XMANAI platform should provide the means to operate and enforce the secure transfer of any data asset (especially data and results) through all platform layers and components (e.g. centralized cloud, private cloud).
Category	III. Data Asset Management and Security (User Journey(s): Business User Phase 1, Data Scientist Phase 2, Data Scientist Phase 3)
Prerequisites	-
Status	**Won't Have**
Changes Since Previous Releases	Updated the Priority "Must-have" (which the feature had in the intermediate MVP release) to Status "Won't-Have": it is expected that data assets will remain in the layer where they were originally stored (i.e. centralized cloud or in the private cloud of a demonstrator) and any necessary processing will be performed on that layer (both for the demonstrator and technical partners).

XMANAI_F_DM_022. Data Asset Access and Activity Logging

Description	The XMANAI platform should keep track of specific activities related to data assets (especially datasets) and monitor the access granted to different organizations.
Category	III. Data Asset Management and Security (User Journey(s): Business User Phase 1, Business User Phase 2, Data Engineer Phase 1, Data Scientist Phase 2, Data Scientist Phase 3)
Prerequisites	XMANAI_F_DM_020
Status	**Will-Have**
Changes Since Previous Releases	Updated the feature description. Updated the Priority "Should-have" (which the feature had in the intermediate MVP release) to Status "Will-Have"

XMANAI_F_DM_023. AI Model Security Assessment

Description	The XMANAI platform should allow for model security assessment by testing the datasets, detecting and filtering possible adversarial data points as well as by increasing the robustness of ML models against adversarial data points.
Category	III. Data Asset Management and Security (User journey(s): Data Scientist Phase 3, Data Engineer Phase 2)
Prerequisites	-
Status	**Will-Have**
Changes Since Previous Releases	Updated the feature description. Updated the Priority "Must-have" (which the feature had in the intermediate MVP release) to Status "Will-Have".

XMANAI_F_DM_024. AI Pipeline Security Assessment

Description	The XMANAI platform should allow for security assessment of the overall AI pipeline, e.g. by testing the training data sets for possible unfair biases.
Category	III. Data Asset Management and Security (User journey(s): Data Scientist Phase 3, Data Engineer Phase 2)
Prerequisites	-
Status	**Won't Have**
Changes Since Previous Releases	Updated the Priority "Should-have" (which the feature had in the intermediate MVP release) to Status "Won't-Have" since security assessment will be provided only at model level (as reflected in the XMANAI_F_DM_023). It is not feasible in the XMANAI project to extend this to the pipeline level (as required in this feature).

XMANAI_F_DS_025. Data Asset Sharing

Description	The XMANAI platform should provide a trustful mechanism for sharing (e.g. datasets, ML/DL models, results) to grant legitimate access to selected organizations. Such a mechanism is based on smart contracts that dictate the terms of the asset exchange/sharing. All users that belong to an organization will have access to a data asset for which data sharing contracts are in place for their organization.
Category	IV. Data Asset Sharing/ Contracts (User journey(s): Business User Phase 2, Data Scientist Phase 3, Data Engineer Phase 2)
Prerequisites	-
Status	Will-Have
Changes Since Previous Releases	Updated the Priority "Must-have" (which the feature had in the intermediate MVP release) to Status "Will-Have".

XMANAI_F_DS_026. Data Asset Trading

Description	The XMANAI platform should provide a contract-based trading mechanism for data assets with selected users/organisations, and support offline, as well as online, payment methods.
Category	IV. Data Asset Sharing/ Contracts (User journey(s): Business User Phase 2, Data Scientist Phase 3, Data Engineer Phase 2)
Prerequisites	-
Status	Won't Have

Changes Since Previous Releases	No change since it was confirmed that the specific feature will not be supported.

XMANAI_F_DS_027. Data Asset Contract Management

Description	The XMANAI platform should provide functionalities related to smart contract management, that include contract preparation, negotiation, approval and enforcement to share data assets (datasets, ML/DL models, XAI pipeline results).
Category	IV. Data Asset Sharing/ Contracts (User journey(s): Business User Phase 2)
Prerequisites	-
Status	Will-Have
Changes Since Previous Releases	Updated the feature description. Updated the Priority "Must-have" (which the feature had in the intermediate MVP release) to Status "Will-Have".

XMANAI_F_DS_028. Data Asset Search and Discovery

Description	The XMANAI platform should allow searching for, discovering and navigating to the different types of data assets (in terms of datasets, ML/DL models, XAI pipeline results). This functionality includes indexing, sorting and filtering data assets, as well as viewing their details.
Category	IV. Data Asset Sharing/ Contracts (User journey(s): Business User Phase 2, Data Scientist Phase 2)
Prerequisites	-
Status	Will-Have
Changes Since Previous Releases	Updated the feature description. Updated the Priority "Must-have" (which the feature had in the intermediate MVP release) to Status "Will-Have".

XMANAI_F_DS_029. Secure Transfer of data assets acquired to their legitimate consumers

Description	The XMANAI platform should provide the appropriate mechanism for the secure transfer of a data asset to a new legitimate data consumer that acquired it and keep the relevant history logs.
Category	IV. Data Asset Sharing/ Contracts (User journey(s): Business User Phase 2)
Prerequisites	XMANAI_F_DM_020
Status	Won't Have
Changes Since Previous Releases	Updated the Priority "Must-have" (which the feature had in the intermediate MVP release) to Status "Won't-Have": Data sharing is expected to be conducted among the demonstrator and technical partners in the context of the XMANAI project. In this context, it is expected that data assets will remain in the layer where they were originally stored (i.e. centralized cloud or in the private cloud of a demonstrator) and any necessary processing will be performed on that layer (both for the demonstrator and technical partners).

XMANAI_F_DP_030. Data View & Visualisation

Description	The XMANAI platform should provide useful data views, including data distribution, raw data sample preview, basic statistics, aggregations, descriptive

	analytics and exploratory queries over the data. In addition, during the XAI pipeline design, a preview of a sample from the execution will be also available in the logs. Once the results of an XAI pipeline are available, appropriate visualizations can be configured by the users with the help of a set of supported charts.
Category	V. Data Preparation (User journey(s): Data Scientist Phase 2, Business User Phase 2)
Prerequisites	XMANAI_F_DI_011
Status	Will-Have
Changes Since Previous Releases	Updated the feature description and related XMANAI components. Updated the Priority "Must-have" (which the feature had in the intermediate MVP release) to Status "Will-Have".

XMANAI_F_DP_031. Data Manipulation

Description	The XMANAI platform should allow for the most common data manipulation functionalities, such as merging, splitting, augmenting, resampling and aggregating data. The creation of new features (columns) and the handling of missing values are also considered to be part of the data manipulation functionalities.
Category	V. Data Preparation (User journey(s): Data Scientist Phase 2)
Prerequisites	XMANAI_F_DI_011
Status	**Will-Have**
Changes Since Previous Releases	Updated the related XMANAI components. Updated the Priority "Must-have" (which the feature had in the intermediate MVP release) to Status "Will-Have".

XMANAI_F_DP_032. Data Transformation

Description	The XMANAI platform should allow for the most common data transformation functionalities, such as normalization, encoding, data type modification, etc.
Category	V. Data Preparation (User journey(s): Data Scientist Phase 2)
Prerequisites	XMANAI_F_DI_011
Status	**Will-Have**
Changes Since Previous Releases	Updated the related XMANAI components. Updated the Priority "Should-have" (which the feature had in the intermediate MVP release) to Status "Will-Have".

XMANAI_F_AI_033. XAI Model Design

Description	The XMANAI platform should allow users to design and configure an AI (ML/DL) model based on a selected algorithm based on different AI/analytics tools/libraries. Such a functionality allows data scientists to import the necessary scripts and to define and configure training and inference sessions.
Category	VI. Data AI Analytics/Pipelines (User journey(s): Data Scientist Phase 3)
Prerequisites	XMANAI_F_DM_017
Status	**Will-Have**
Changes Since Previous Releases	Updated the feature title and description. Updated the Priority "Must-have" (which the feature had in the intermediate MVP release) to Status "Will-Have".

XMANAI_F_AI_034. XAI Model Training, Application & Evaluation

Description	The XMANAI platform should provide the appropriate mechanisms for training, application and evaluation of AI (ML/DL) models through the configuration of appropriate sessions and their inclusion in XAI pipelines. This functionality overall includes parameter configuration and optimization, evaluation metrics definition and monitoring. The differences between the experimentation stage and the production stage should be also considered during the process.
Category	VI. Data AI Analytics/Pipelines (User journey(s): Data Scientist Phase 3)
Prerequisites	XMANAI_F_AI_033
Status	**Will-Have**
Changes Since Previous Releases	Updated the feature title and description. Updated the Priority “Must-have” (which the feature had in the intermediate MVP release) to Status “Will-Have”.

XMANAI_F_AI_035. XAI Pipeline Design

Description	The XMANAI platform should provide appropriate AI pipeline design functionalities, where a complete pipeline can be defined (as workflow) and each of its ingredients (tasks) can be configured. It should also anticipate the addition of annotations / comments, the selection of appropriate execution environment (in the centralized cloud or the selected stakeholder’s infrastructures) and the definition of the execution schedule.
Category	VI. Data AI Analytics/Pipelines (User journey(s): Data Scientist Phase 3, Data Engineer Phase 2)
Prerequisites	XMANAI_F_DM_018
Status	**Will-Have**
Changes Since Previous Releases	Updated the feature title and description, as well as the related XMANAI components. Updated the Priority “Must-have” (which the feature had in the intermediate MVP release) to Status “Will-Have”.

XMANAI_F_AI_036. XAI Pipeline Execution & Evaluation on XMANAI Common Cloud

Description	The XMANAI platform should provide the appropriate centralized cloud execution environment for XAI pipeline execution and evaluation. This includes the overall run (once or based on the schedule set), evaluation, storage of models and results, calculation of the associated metrics and storage of execution logs.
Category	VI. Data AI Analytics/Pipelines (User journey(s): Data Scientist Phase 5, Data Engineer Phase 3)
Prerequisites	XMANAI_F_DM_035
Status	**Will-Have**
Changes Since Previous Releases	Updated the feature title and description. Updated the Priority “Must-have” (which the feature had in the intermediate MVP release) to Status “Will-Have”.

XMANAI_F_AI_037. XAI Pipeline Execution & Evaluation on Premise / Private Cloud

Description	The execution and evaluation of an AI pipeline should be also available to run in a private cloud or in on-premise infrastructures in general, that are provided by each

	demonstrator, either during experimentation or production. The configuration of this process works similarly to the centralized cloud execution.
Category	VI. Data AI Analytics/Pipelines (User journey(s): Business User Phase 2, Data Scientist Phase 5, Data Engineer Phase 3)
Prerequisites	XMANAI_F_DM_035
Status	**Will-Have**
Changes Since Previous Releases	Updated the feature title and description. Updated the Priority "Must-have" (which the feature had in the intermediate MVP release) to Status "Will-Have".

XMANAI_F_AI_038. Collaboration over XAI pipelines creation

Description	The XMANAI platform should allow collaboration among different users during the configuration of XAI pipelines by allowing them to create comments/annotations in each task, replying to other's comments and resolving comments. In addition, the collaboration capabilities at the XAI pipeline level extend to the comparison of different experiments and models' performance. XMANAI will maintain a history of the various events and provide a simulation environment for different settings for the same model.
Category	VI. Data AI Analytics/Pipelines (User journey(s): Data Scientist Phase 3, Data Engineer Phase 2)
Prerequisites	XMANAI_F_DM_035
Status	**Will-Have**
Changes Since Previous Releases	Updated the feature title and description. Updated the Priority "Must-have" (which the feature had in the intermediate MVP release) to Status "Will-Have".

XMANAI_F_AI_039. XAI Pipeline Results Management

Description	The XMANAI platform should support management of the XAI pipeline results, allowing users to store the generated results (on the location where the XAI pipeline was executed), view their summary statistics and a sample (in the execution logs), and retrieve them via API.
Category	VI. Data AI Analytics/Pipelines (User journey(s): Data Scientist Phase 4)
Prerequisites	XMANAI_F_DM_037
Status	**Will-Have**
Changes Since Previous Releases	Updated the feature title and description. Updated the Priority "Must-have" (which the feature had in the intermediate MVP release) to Status "Will-Have".

XMANAI_F_AI_040. XAI Pipeline Results Visualization

Description	The XMANAI platform should allow users to visualize the results of an XAI pipeline by selecting and configuring a single or multiple charts and graphs among a list of various supported charts and graphs. Saving or exporting the visualisations are also part of this feature.
Category	VI. Data AI Analytics/Pipelines (User journey(s): Business User Phase 3, Data Scientist Phase 4)
Prerequisites	XMANAI_F_DM_037

Status	**Will-Have**
Changes Since Previous Releases	Updated the feature title and description. Updated the Priority "Must-have" (which the feature had in the intermediate MVP release) to Status "Will-Have".

XMANAI_F_AI_041. XAI Pipeline Support

Description	The XMANAI platform should provide a set of supporting functionalities related to the AI pipeline execution in the centralized cloud or the selected stakeholder's "on-premise" environment, which include, but are not limited to, viewing the execution logs and handling errors for each step of each occurred execution.
Category	VI. Data AI Analytics/Pipelines (User journey(s): Data Scientist Phase 3, Data Engineer Phase 2)
Prerequisites	XMANAI_F_DM_035, XMANAI_F_DM_037
Status	**Will-Have**
Changes Since Previous Releases	Updated the feature title and description. Updated the Priority "Could-have" (which the feature had in the intermediate MVP release) to Status "Will-Have".

XMANAI_F_EX_042. Explainability Methods Management

Description	The XMANAI platform should effectively manage explainability methods, in terms of adding, removing and configuring different explainability techniques in training/inference sessions in order to be eventually included in XAI pipelines.
Category	VII. AI Model/Results Explainability (User journey(s): Business User Phase 3, Data Scientist Phase 4)
Prerequisites	XMANAI_F_AI_033, XMANAI_F_AI_035
Status	**Will-Have**
Changes Since Previous Releases	Updated the feature title and description. Updated the Priority "Must-have" (which the feature had in the intermediate MVP release) to Status "Will-Have".

XMANAI_F_EX_043. Collaboration over AI explanations

Description	The XMANAI platform should support the cooperation of different stakeholders on the application of explainability methods at AI model or pipeline level. Viewing the results/explanations, requesting for more details in an already provided explanation and adding notes/comments are key functionalities to ensure that explanations will be eventually appropriate for use by the target stakeholders (business users).
Category	VII. AI Model/Results Explainability (User journey(s): Business User Phase 3, Data Scientist Phase 4)
Prerequisites	XMANAI_F_AI_033, XMANAI_F_AI_035, XMANAI_F_AI_033, XMANAI_F_EX_042, XMANAI_F_EX_044
Status	**Won't-Have**
Changes Since Previous Releases	Updated the feature description. Updated the Priority "Must-have" (which the feature had in the intermediate MVP release) to Status "Won't-Have" as the implementation of such a collaboration functionality at explanation level requires a level of interaction that is not feasible (time and effort-wise) to be supported in XMANAI (in the phase the project is currently).

XMANAI_F_EX_044. XAI Results & Explanations Visualisation

Description	The XMANAI platform should provide comprehensible visualisations of the results and explanations of XAI models, through custom charts and graphs (supported by the different explainability techniques or by popular visualization libraries), adjusted in accordance with the user profile (i.e. business user vs. data scientist)
Category	VII. AI Model/Results Explainability (User journey(s): Business User Phase 3, Data Scientist Phase 4)
Prerequisites	XMANAI_F_AI_033, XMANAI_F_AI_035, XMANAI_F_AI_033, XMANAI_F_EX_042
Status	**Will-Have**
Changes Since Previous Releases	Updated the feature title and description. Updated the Priority "Must-have" (which the feature had in the intermediate MVP release) to Status "Will-Have".

XMANAI_F_EX_045. XAI Results Evaluation

Description	The XMANAI platform should allow for manual feedback and validation of the explainability results by the targeted users in order to improve the explanations provided per AI pipeline or model.
Category	VII. AI Model/Results Explainability (User journey(s): Business User Phase 3)
Prerequisites	XMANAI_F_AI_033, XMANAI_F_AI_035, XMANAI_F_AI_033, XMANAI_F_EX_042, XMANAI_F_EX_044
Status	**Will-Have**
Changes Since Previous Releases	Updated the feature title and description. Updated the Priority "Must-have" (which the feature had in the intermediate MVP release) to Status "Will-Have".

4.3 Final MVP Updates Summary

A summary of the status of all features along their assessment in the intermediate MVP and the final MVP is provided in the following table.

Table 4-1: Final XMANAI MVP Consolidation

MVP Feature ID	MVP Feature Title	Preliminary MVP Assessment	Final MVP Assessment
XMANAI_F_UM_001	Organization-based access (users within the same organization can access all data assets that belong to an organization)	Must-have	Will-Have
XMANAI_F_UM_002	Project-based access (users from different organizations can access data assets that are grouped under a project)	Won't-have	Won't-Have
XMANAI_F_UM_003	Delete all data assets	Could-have	Will-Have
XMANAI_F_DI_004	Data Sources Management (add/edit/remove/configure/schedule)	Must-have	Will-Have
XMANAI_F_DI_005	Data Secure Uploading as file(s)	Must-have	Will-Have
XMANAI_F_DI_006	Data Secure Uploading via API	Must-have	Will-Have
XMANAI_F_DI_007	Data Mapping to a Data Model and Harmonization (common semantics, data type casting, ~~transformation to common measurement unit/timestamp~~)	Should-have	Will-Have
XMANAI_F_DI_008	Data Model Management (view model, add/edit/remove model concepts, ~~versioning different representations~~, proposing changes)	Must-have	Will-Have
XMANAI_F_DI_009	Data Cleansing (quality checks, cleaning rules definition and execution before storage)	Won't-have	Won't-Have

MVP Feature ID	MVP Feature Title	Preliminary MVP Assessment	Final MVP Assessment
XMANAI_F_DI_010	Data Anonymisation (configuration and execution before making them available in XMANAI)	Could-have	Will-Have
XMANAI_F_DI_011	Data Storage in central XMANAI Cloud Storage	Must-have	Will-Have
XMANAI_F_DI_012	Data Storage on-premise / in private cloud	Must-have	Will-Have
XMANAI_F_DI_013	Data Encryption (in transfer, in storage)	Won't-have	Won't-Have
XMANAI_F_DM_014	Dataset Management (add/edit/remove asset and metadata, versioning)	Must-have	Will-Have
XMANAI_F_DM_015	Features Management (store, add/edit/remove asset and metadata, define rules, versioning)	Won't-have	Won't-Have
XMANAI_F_DM_016	Results Management (edit/remove asset and metadata, versioning)	Should-have	Will-Have
XMANAI_F_DM_017	AI Model Management (create/store/update/delete, configure, ~~versioning, register/import~~)	Must-have	Will-Have
XMANAI_F_DM_018	AI Pipeline Management (create/store/update/delete/~~clone~~, view ~~IPR of~~ assets involved (depending on their IPR), ~~define templates, versioning~~)	Must-have	Will-Have
XMANAI_F_DM_019	Data Asset Export (download file or via API depending on IPR)	Should-have	Will-Have
XMANAI_F_DM_020	Data Asset Access Policies and Licencing (configuration, management and enforcing)	Must-have	Will-Have
XMANAI_F_DM_021	Data Asset Secure Transfer (through platform layers, and components)	Must-have	Won't-Have
XMANAI_F_DM_022	Data Asset Access and Activity Logging	Should-have	Will-Have
XMANAI_F_DM_023	AI Model Security Assessment	Must-have	Will-Have
XMANAI_F_DM_024	AI Pipeline Security Assessment	Should-have	Won't-Have
XMANAI_F_DS_025	Data Asset Sharing (based on contracts, with selected ~~users /~~ organizations)	Must-have	Will-Have
XMANAI_F_DS_026	Data Asset Trading (based on contracts, payment performed offline/ online)	Won't-have	Won't-Have
XMANAI_F_DS_027	Data Asset Contract Management (contract preparation, negotiation, agreement, enforcement)	Must-have	Will-Have
XMANAI_F_DS_028	Data Asset Search and Discovery (including indexing, sorting, filtering, ~~matching level to your asset,~~ view details)	Should-have	Will-Have
XMANAI_F_DS_029	Secure Transfer of data assets acquired to their legitimate consumers (across platform layers, history log)	Should-have	Won't-Have
XMANAI_F_DP_030	Data View & Visualisation (query data, view distribution, statistics, aggregations, time periods, descriptive analytics, preview sample)	Must-have	Will-Have
XMANAI_F_DP_031	Data Manipulation (merge, split, augment, resample, aggregate, create new features, handle missing values, etc.)	Must-have	Will-Have
XMANAI_F_DP_032	Data Transformation (normalisation, encoding, modifying data types, etc.)	Should-have	Will-Have
XMANAI_F_AI_033	AI Model Design (define, configure, store, import script, ~~export~~)	Must-have	Will-Have
XMANAI_F_AI_034	AI Model Training, Application & Evaluation (experimentation vs production, configure control parameters, define/monitor eval. metrics, ~~save check points,~~ support parameter optimisation)	Must-have	Will-Have
XMANAI_F_AI_035	AI Pipeline Design (define, configure~~, register~~ AI model, add annotations/comments, reuse common features)	Must-have	Will-Have
XMANAI_F_AI_036	AI Pipeline Execution & Evaluation on XMANAI Common Cloud (experimentation vs production, ~~run automatic tests for AI models,~~ run on scheduling, configuration)	Must-have	Will-Have
XMANAI_F_AI_037	AI Pipeline Execution & Evaluation on Premise / Private Cloud (experimentation vs production, ~~run automatic tests for AI models,~~ run on scheduling, configuration)	Must-have	Will-Have

MVP Feature ID	MVP Feature Title	Preliminary MVP Assessment	Final MVP Assessment
XMANAI_F_AI_038	Collaboration over AI pipelines creation (experiments comparison, history of events, simulations of different settings~~, models, and methods~~ for same task/model)	Must-have	Will-Have
XMANAI_F_AI_039	AI Pipeline Results Management (store, export, add summary statistics, ~~easily compare with real values (for predictions),~~ retrieve via API)	Must-have	Will-Have
XMANAI_F_AI_040	AI Pipeline Results Visualisation (configuration of various charts, add comments, store, export, run on cloud vs on premise)	Must-have	Will-Have
XMANAI_F_AI_041	AI Pipeline Support (common metadata model, execution logs, error handling, ~~show comp. resources expended~~)	Could-have	Will-Have
XMANAI_F_EX_042	Explainability Methods Management (add/remove/configure, register/import)	Must-have	Will-Have
XMANAI_F_EX_043	Collaboration over AI model/results/pipelines explanations (application of explainability methods at AI pipeline or model level, results querying)	Must-have	Won't-Have
XMANAI_F_EX_044	XAI Results & Explanations Visualisation (various charts, adjust based on user profile)	Must-have	Will-Have
XMANAI_F_EXI_045	Explainability Results Evaluation (allow manual feedback & results validation)	Must-have	Will-Have

4.4 MVP & Technical Requirements Prioritisation

The technical requirements that are relevant per feature of the XMANAI MVP along with their priority are presented in the following table.

Table 4-2: Final XMANAI MVP & Technical Requirements Consolidation

MVP Feature ID	MVP Feature Title	Related Requirement & Prioritization	Final MVP Assessment
XMANAI_F_UM_001	Organization-based access	TR_135 - High	Will-Have
XMANAI_F_UM_002	Project-based access	-	Won't-Have
XMANAI_F_UM_003	Delete all data assets	TR_4 - High	Will-Have
XMANAI_F_DI_004	Data Sources Management	TR_1 – High, TR_2 – Medium, TR_16 - High	Will-Have
XMANAI_F_DI_005	Data Secure Uploading as file(s)	TR_3 - High	Will-Have
XMANAI_F_DI_006	Data Secure Uploading via API	TR_1, T_2, TR_15, TR_16 - Medium	Will-Have
XMANAI_F_DI_007	Data Mapping to a Data Model and Harmonization	TR_13 – High, TR_54 - Deprecated	Will-Have
XMANAI_F_DI_008	Data Model Management	TR_5 - Medium, TR_6 - Medium, TR_7 - Medium, TR_8 - High, TR_9 - Low, TR_10 - Medium, TR_11 - Low	Will-Have
XMANAI_F_DI_009	Data Cleansing (quality checks, cleaning rules definition and execution before storage)	TR_14 - Deprecated	Won't-Have
XMANAI_F_DI_010	Data Anonymisation	TR_29 - Medium, TR_30 - Medium	Will-Have
XMANAI_F_DI_011	Data Storage in central XMANAI Cloud Storage	TR_3	Will-Have
XMANAI_F_DI_012	Data Storage on-premise / in private cloud	TR_20 - High	Will-Have
XMANAI_F_DI_013	Data Encryption (in transfer, in storage)	-	Won't-Have
XMANAI_F_DM_014	Dataset Management	TR_12 – High, TR_15 – High, TR_42 – Medium, TR_43 – Medium, TR_93 - Medium	Will-Have
XMANAI_F_DM_015	Features Management	TR_74 – Deprecated, TR_76 - Deprecated	Won't-Have

MVP Feature ID	MVP Feature Title	Related Requirement & Prioritization	Final MVP Assessment
XMANAI_F_DM_016	Results Management	TR_44 – Medium, TR_81 – Low, TR_93 - Medium	Will-Have
XMANAI_F_DM_017	AI Model Management	TR_44 – Medium, TR_93 – Medium, TR_118 – Medium, TR_119 - Medium, TR_120 - Medium, TR_121 - Medium	Will-Have
XMANAI_F_DM_018	AI Pipeline Management	TR_49 – Medium, TR_70 – Deprecated, TR_71 – Deprecated,	Will-Have
XMANAI_F_DM_019	Data Asset Export (download file or via API depending on IPR)	TR_17 - High, TR_18 – High, TR_19 - Medium	Will-Have
XMANAI_F_DM_020	Data Asset Access Policies and Licencing (configuration, management and enforcing)	TR_21 - High, TR_22 - High, TR_23 - High, TR_24 - High, TR_25 - High, TR_26 – Deprecated, TR_27 - High, TR_28 - High	Will-Have
XMANAI_F_DM_021	Data Asset Secure Transfer (through platform layers, and components)	TR_31 - Low	Won't-Have
XMANAI_F_DM_022	Data Asset Access and Activity Logging	TR_45 - Medium, TR_46 - Medium, TR_47 - Medium, TR_48 - Medium	Will-Have
XMANAI_F_DM_023	AI Model Security Assessment	TR_131 - Medium, TR_132 – Medium, TR_133 – Deprecated, TR_134 - Deprecated	Will-Have
XMANAI_F_DM_024	AI Pipeline Security Assessment	-	Won't-Have
XMANAI_F_DS_025	Data Asset Sharing	TR_32 – Medium, TR_37 - Medium	Will-Have
XMANAI_F_DS_026	Data Asset Trading	TR_33 - Deprecated, TR_38- Deprecated	Won't-Have
XMANAI_F_DS_027	Data Asset Contract Management	TR_39 - Medium, TR_40 - Medium	Will-Have
XMANAI_F_DS_028	Data Asset Search and Discovery	TR_34 – Medium, TR_35 – Medium, TR_36 – Medium, TR_41 - Medium	Will-Have
XMANAI_F_DS_029	Secure Transfer of data assets acquired to their legitimate consumers (across platform layers, history log)	-	Won't-Have
XMANAI_F_DP_030	Data View & Visualisation	TR_50 - Medium, TR_51 - Medium, TR_52 - Medium, TR_53 - Medium, TR_54 – Deprecated, TR_63 – Medium, TR_64 - Medium	Will-Have
XMANAI_F_DP_031	Data Manipulation	TR_55 - Medium, TR_56 - Medium, TR_59 - Medium, TR_62 - Medium	Will-Have
XMANAI_F_DP_032	Data Transformation	TR_57 - Medium, TR_58 - Medium, TR_60 - Medium, TR_61 - Medium	Will-Have
XMANAI_F_AI_033	AI Model Design	TR_94 – Medium, TR_110 – High, TR_124 - Medium	Will-Have
XMANAI_F_AI_034	AI Model Training, Application & Evaluation	TR_75 – Deprecated, TR_85 – Medium, TR_87 – Medium, TR_95 – Low, TR_97 - Medium, TR_98 - Medium, TR_99 - Medium, TR_100 - Medium, TR_103 – Medium, TR_125 - High, TR_126 – Medium, TR_128 – Medium, TR_129 – Low, TR_130 - High	Will-Have
XMANAI_F_AI_035	AI Pipeline Design	TR_65 - Medium, TR_66 - Medium, TR_67 - Medium, TR_68 - Medium, TR_72 – Deprecated, TR_79 - Medium, TR_80 - Medium, TR_83 – Medium, TR_85, TR_96 - Medium, TR_114, TR_115, TR_116	Will-Have
XMANAI_F_AI_036	AI Pipeline Execution & Evaluation on XMANAI Common Cloud	TR_73 – Deprecated, TR_86 - Medium, TR_104 – High, TR_105 – Medium	Will-Have

MVP Feature ID	MVP Feature Title	Related Requirement & Prioritization	Final MVP Assessment
XMANAI_F_AI_037	AI Pipeline Execution & Evaluation on Premise / Private Cloud	TR_73 – Deprecated, TR_86 – Medium, TR_104 – High, TR_105 – Medium, TR_108 – Medium, TR_113 - High	Will-Have
XMANAI_F_AI_038	Collaboration over AI pipelines creation	TR_69 - Medium, TR_103, TR_126, TR_128, TR_129	Will-Have
XMANAI_F_AI_039	AI Pipeline Results Management	TR_81 – Medium, TR_82 – Medium, TR_110 - High, TR_111 - High	Will-Have
XMANAI_F_AI_040	AI Pipeline Results Visualisation	TR_88 - Medium, TR_89 - High, TR_90 - Medium, TR_113	Will-Have
XMANAI_F_AI_041	AI Pipeline Support	TR_76 – Deprecated, TR_93 – Medium, TR_101 - Medium, TR_102 – Medium, TR_106 - Deprecated, TR_107 – Medium, TR_109 – Medium, TR_112 - Medium	Will-Have
XMANAI_F_EX_042	Explainability Methods Management	TR_77 - High, TR_78 – Medium, TR_91 – High, TR_117 – Medium, TR_122 - Medium, TR_123 - Medium	Will-Have
XMANAI_F_EX_043	Collaboration over AI model/results/pipelines explanations	TR_92 - Deprecated	Won't-Have
XMANAI_F_EX_044	Explainability Results & Explanations Visualisation	TR_84 - Medium	Will-Have
XMANAI_F_EX_045	Explainability Results Evaluation	TR_127 -High	Will-Have

4.5 MVP vs XMANAI Components Functionalities

The MVP features that are eventually supported in XMANAI are realised by the different components developed in WP2-WP4, as depicted in the following table.

Table 4-3: Final XMANAI MVP & Platform Components Consolidation

MVP Feature ID	MVP Feature Title	Related Components
XMANAI_F_UM_001	Organization-based access	• Identity & Authorization Management
XMANAI_F_UM_003	Delete all data assets	• Data Handler: File/Data Manager, • Data Handler: Data Gateway, • XAI Marketplace: Registry/Metadata Manager
XMANAI_F_DI_004	Data Sources Management	• Data Handler: API Data Harvester
XMANAI_F_DI_005	Data Secure Uploading as file(s)	• Data Handler: File/Data Manager
XMANAI_F_DI_006	Data Secure Uploading via API	• Data Handler: API Data Harvester, • Data Storage Services: Assets Store
XMANAI_F_DI_007	Data Mapping to a Data Model and Harmonization	• Data Handler: File Data Harvester • Data Handler: API Data Harvester • Knowledge Graph Manager
XMANAI_F_DI_008	Data Model Management	• Knowledge Graph Manager
XMANAI_F_DI_010	Data Anonymisation	• Anonymiser
XMANAI_F_DI_011	Data Storage in central XMANAI Cloud Storage	• Data Storage Services: Assets Store with Version Control
XMANAI_F_DI_012	Data Storage on-premise / in private cloud	• Data Storage Services: Assets Store with Version Control available through an On-Premise Environment installation
XMANAI_F_DM_014	Dataset Management	• Data Handler: File Data Harvester • Data Handler: File/Data Manager

MVP Feature ID	MVP Feature Title	Related Components
		• XAI Marketplace: Registry/Metadata Manager • Provenance Engine • Data Storage Services: Assets Store with Version Control
XMANAI_F_DM_016	Results Management	• Execution & Orchestration Engine • XAI Marketplace: Registry/Metadata Manager • Provenance Engine, • Data Storage Services: Assets Store with Version Control
XMANAI_F_DM_017	AI Model Management	• XAI Model Engineering Engine • Experiment Tracking Engine
XMANAI_F_DM_018	AI Pipeline Management	• Pipeline Designer
XMANAI_F_DM_019	Data Asset Export	• Data Handler: Data Exporter • Data Storage Services: Assets Store with Version Control
XMANAI_F_DM_020	Data Asset Access Policies and Licencing	• Access Manager: Policy Engine, • Access Manager: Policy Editor, • XAI Marketplace: Contract Manager
XMANAI_F_DM_021	Data Asset Secure Transfer	• Execution & Orchestration Engine
XMANAI_F_DM_022	Data Asset Access and Activity Logging	• Access Manager: Policy Engine, • Provenance Engine
XMANAI_F_DM_023	AI Model Security Assessment	• XAI Model Guard
XMANAI_F_DS_025	Data Asset Sharing	• XAI Marketplace: Registry/Metadata Manager, • Access Manager: Policy Engine, • Identity & Authorisation Management
XMANAI_F_DS_027	Data Asset Contract Management	• XAI Marketplace: Contract Manager
XMANAI_F_DS_028	Data Asset Search and Discovery	• XAI Marketplace: Registry/Metadata Manager • Access Manager: Policy Engine
XMANAI_F_DP_030	Data View & Visualisation	• Interactive Data Exploration & Experimentation Tool, • Data Preparation Engine, Results Visualization Engine, • Pipeline Serving & Monitoring Engine
XMANAI_F_DP_031	Data Manipulation	• Data Preparation Engine • Interactive Data Exploration & Experimentation Tool
XMANAI_F_DP_032	Data Transformation	• Data Preparation Engine, • Model Engineering Engine
XMANAI_F_AI_033	AI Model Design	• Model Engineering Engine
XMANAI_F_AI_034	AI Model Training, Application & Evaluation	• Model Engineering Engine, • Pipeline Designer, • Execution & Orchestration Engine
XMANAI_F_AI_035	AI Pipeline Design	• Pipeline Designer, • Pipeline Serving & Monitoring Engine
XMANAI_F_AI_036	AI Pipeline Execution & Evaluation on XMANAI Common Cloud	• Pipeline Serving & Monitoring Engine, • Execution & Orchestration Engine
XMANAI_F_AI_037	AI Pipeline Execution & Evaluation on Premise / Private Cloud	• Pipeline Serving & Monitoring Engine, • Execution & Orchestration Engine
XMANAI_F_AI_038	Collaboration over AI pipelines creation	• Pipeline Designer, • Experiment Tracking Engine
XMANAI_F_AI_039	AI Pipeline Results Management	• Pipeline Serving & Monitoring Engine, • Execution & Orchestration Engine, • Data Storage Services: Assets Store with Version Control
XMANAI_F_AI_040	AI Pipeline Results Visualisation	• Results Visualization Engine
XMANAI_F_AI_041	AI Pipeline Support	• Pipeline Serving & Monitoring Engine

MVP Feature ID	MVP Feature Title	Related Components
XMANAI_F_EX_042	Explainability Methods Management	• XAI Model Explanations Engine, • Model Engineering Engine
XMANAI_F_EX_044	Explainability Results & Explanations Visualisation	• XAI Model Explanations Engine, • Results Visualization Engine, • Execution & Orchestration
XMANAI_F_EX_045	Explainability Results Evaluation	• XAI Model Explanations Engine, • Execution & Orchestration

5 XMANAI Concept in a Nutshell

Taking into consideration the final MVP as described in Section 4, the concept of X-By-Design (Explainable-by-Design) that XMANAI brings forward has been refined to consider the research and development priorities that have been set.

Based on a user-driven approach involving all stakeholders (data scientists, data engineers, business users), XMANAI aims to render humans capable of fully understanding how decisions are being made step-by-step through customized explanations that are created and served in different ways (e.g. text, visual charts) to facilitate learning and answering different types of human-oriented (what/how/why) explanatory questions, that currently remain unanswered, such as: (a) Why did the AI system make this specific Prediction A? (b) Why didn't the AI system make Prediction B for the same circumstances? (c) Under what circumstances would the AI system make prediction C? (d) Why does the AI system provide the prediction A in the next hour and prediction B in 4 hours?

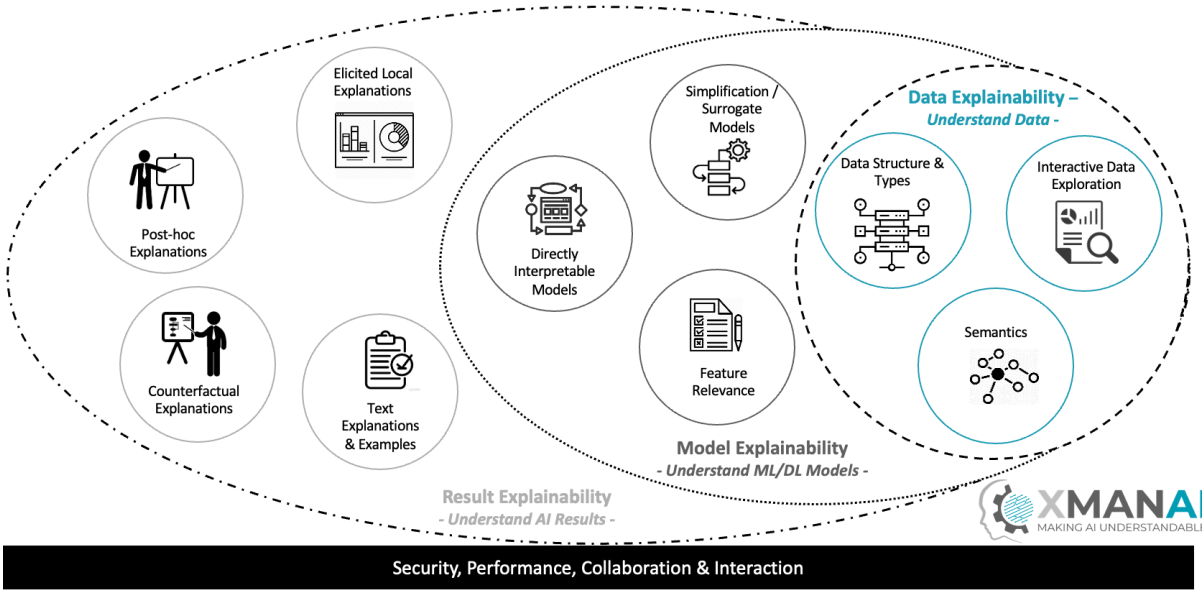

Figure 5-1: XMANAI X-By-Design approach from a technical perspective

In order to transition towards an Explainable-by-Design paradigm from a technical perspective, XMANAI views and supports explainability under three (3) interlinked and collaborative perspectives:

- **Data Explainability** that focuses on a concrete **understanding of data** in terms of semantics and structure (data types) per feature in order to gain insights into the input data, achieved through mapping to a common data model. Interactive data exploration and visualization that allow viewing data distribution/profiling charts or summary statistics (e.g. number of missing values, min/max values) can be leveraged in order to monitor potential data drifting or bias issues. XMANAI contributes to this axis through:
- **Model Explainability** that concerns **understanding the different AI (ML/DL) models** towards *global* interpretability (answering how does a model work for all our predictions) and *local* interpretability (answering how a model is generating a specific prediction, given specific data points). Different explainability techniques can be employed depending on the model family and type, e.g. black-box or opaque models, in order to attempt to shed light on the model design and training phase. Indicative categories include: (a) Explanations by simplification (or surrogate models) referring to the techniques that approximate an opaque model using a simpler one, which is easier to interpret; (b) Feature relevance explanations which attempt to explain a model's decision by quantifying the influence of each input variable, serving as a first step towards gaining insights into the model's reasoning; (c) Explanations through directly interpretable models (since

transparent models like decision trees, linear and logistic regression are directly interpretable). Typical techniques associated to the above categories are: Local interpretable model-agnostic explanations (LIME), SHAP (SHapley Additive exPlanation), Anchors (High-Precision Model-Agnostic Explanations), Partial Dependence Plot (PDP), Counterfactual Explanations, CAMEL (Causal Models to Explain Learning).

- **Result Explainability** that promotes **shared understanding of results** and translating them into concrete actions in an appropriate style/format. At this step, post-hoc explanations over the results are created (after the model is trained) and may include: (a) Visualizations that typically act as the primer for communicating results to the involved stakeholders in order to inform them about the decision boundary or how features interact with each other; (b) Text explanations that convey in natural language how to take action and can be automatically generated (through natural language generation techniques) or elicited with the involvement of the target audiences; (c) Explanations by example that extract representative instances from the training dataset to demonstrate how the model operates in a similar manner as humans may approach explanations by providing specific examples to describe a broader outcome/process; (d) Counterfactual Explanations that aim to find the model's decision boundaries with respect to specific input values.

In order to produce AI models and pipelines that are explainable by design, XMANAI shall deliver appropriate assessment methods and techniques to address a number of complementary challenges that currently constitute significant data scientists' pains, and are classified under:

- XAI Models Security that performs a risk and vulnerability assessment over different Explainable ML/DL models to offer immunity and robustness. The aim is to timely anticipate and safeguard them against unintended bias and adversarial attacks that may try to manipulate their underlying algorithm after learning what input should be fed so as to lead it to a specific output. Such adversarial attacks may refer to poisoning (attempting to maliciously manipulate the training dataset) and evasion attacks, in general.
- XAI Models Performance, referring to: (a) the technical metrics over the performance of an Explainable ML/DL model for which baseline thresholds need to be set in order to be tracked in parallel with transparency and security. Typical performance metrics are accuracy, F1, sensitivity, scalability; (b) the pure explainability aspects (e.g. explanations usefulness, satisfaction, fidelity, completeness, ambiguity, lack of bias) to gauge how appropriate (in terms of content and delivery method) the generated explanations have been as evaluated by the target audiences that consumed the explanations (rather than the experts who created the corresponding XAI models).
- XAI Assets Sharing that allows interaction among different types of stakeholders who share their assets and collaboratively work to build a solution for a specific manufacturing case/problem that reaches consensus and is broadly accepted.

6 Conclusions and Next Steps

The present deliverable (D1.4) documents the results of the final iteration of tasks T1.3 and T1.4 which have three main objectives: to elaborate on the XMANAI concept (what does XMANAI bring to different stakeholders), to define a set of up-to-date technical requirements, and to provide the final version of the MVP (Minimum Viable Product) that has guided the development and integration activities since their very beginning. To derive these outcomes, a clear and easily comprehensive approach was followed, including: (a) update of the user journeys containing TO-BE scenarios for business users, data scientists and data engineers, (b) maintenance of the technical requirements backlog and organization into features, (c) consolidation of the final MVP.

The technical requirements had been originally based on brainstorming in the Miro boards connected to the user journeys (in D1.2) yet their backlog was continuously monitored and refined (in D1.3 and here, D1.4) to ensure it follows the project development activities. By grouping and organizing such requirements in the backlog, their complementarity with the business requirements and the different steps of the user journeys was revisited.

Reflecting the experience from the development of the Data & AI Services Bundles in WP2 “Industrial Asset Management and Secure Asset Sharing Bundles” and WP3 “Core Artificial Intelligence Bundles for Algorithm Lifecycle Management”, and the AI Models-related tasks of WP4 “Novel Artificial Intelligence Algorithms for Industrial Data Insights Generation”, the final MVP has been defined and agreed among partners. Instead of designating a set of “must-have”, “should-have”, “could-have” and “won't-have right now” features, the final MVP leverages the project’s (technical/platform and demonstration/end-users) advancements to-date to present a clearer positioning with “Will-Have” and “Won’t Have” features. It needs to be noted once again that the XMANAI MVP represents not only a set of features to implement, but also the mentality of work that ensures the XMANAI platform will be appropriately validated by its stakeholders and will deliver the maximum added value, with the lowest possible risk.

The “X-By-Design” concept that is promoted by XMANAI has been also visited from a technical perspective in order to elaborate on the data, model and result viewpoints and their repercussions and advantages.

Finally, D1.4 represents the culmination of the work in WP1 and has further reflected on, improved and concluded its previous editions, namely D1.2 and D1.3.

References

Gerdeman, D., 2017. Companies love big data but lack the strategy to use it effectively. Harv Bus Sch Work Knowl.

Ismail, A., Truong, H.L. and Kastner, W., 2019. Manufacturing process data analysis pipelines: a requirements analysis and survey. Journal of Big Data, 6(1), pp.1-26.

Lenz, J., Wuest, T. and Westkämper, E., 2018. Holistic approach to machine tool data analytics. Journal of manufacturing systems, 48, pp.180-191.

Luke Posey, "Engineering + Data Science: The Missing Duo", 2019. Available at: https://towardsdatascience.com/engineering-data-science-the-ultimate-yet-somehow-missing-duo-597eb21dda98

Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Flach, P., Hernández-Orallo, J., Kull, M., Lachiche, N. and Ramírez-Quintana, M.J., 2017. Casp-dm: Context aware standard process for data mining. arXiv preprint arXiv:1709.09003.

Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Orallo, J.H., Kull, M., Lachiche, N., Quintana, M.J.R. and Flach, P.A., 2019. CRISP-DM twenty years later: From data mining processes to data science trajectories. IEEE Transactions on Knowledge and Data Engineering.

Ries, E., 2009. Venture Hacks interview:" What is the minimum viable product?". URl: http://www. startuplessonslearned. com/2009/03/minimum-viable-product. html(visited on 17/06/2017).

Schmidt, C. and Sun, W.N., 2018. Synthesizing agile and knowledge discovery: case study results. Journal of Computer Information Systems, 58(2), pp.142-150.

XMANAI Description of Action (DoA), 2020.

XMANAI Deliverable D1.1 "State-of-the Art Review in XMANAI Research Domains", 2021.

XMANAI Deliverable D1.2 "XMANAI Concept Detailing, Initial Requirements, Usage Scenarios and Draft MVP", 2021.

XMANAI Deliverable D1.3 "Updated Requirements and AI/Graph Analytics focused MVP", 2022.

XMANAI Deliverable D2.1 "Asset Management Bundles Methods and System Designs", 2021.

XMANAI Deliverable D2.2 "XMANAI Asset Management Bundles- First Release", 2022.

XMANAI Deliverable D3.1 "AI Bundles Methods and System Designs", 2021.

XMANAI Deliverable D3.2 "XMANAI AI Bundles - First Release", 2022.

XMANAI Deliverable D4.1 "Draft Catalogue of XMANAI AI and Graph Machine Learning Models", 2022.

XMANAI Deliverable D4.2 "XMANAI Al and Graph Machine Learning Models – First Round", 2022.

XMANAI Deliverable D5.1 "System Architecture, Bundles Placement Plan and APIs Design", 2021.

XMANAI Deliverable D5.2 "XMANAI Platform- Alpha Version", 2022.

XMANAI Deliverable D6.1 "Demonstrators Requirements", 2021.

XMANAI Deliverable D6.3 "Updated Demonstrators Requirements and New Models and Algorithm Requests", 2023.

Wake, B., 2003. Invest in good stories and smart tasks, August 2003. Retrieved from: http://xp123. com/articles/invest-in-good-stories-and-smart-tasks.

List of Acronyms/Abbreviations

Acronym/ Abbreviation	Description
AI	Artificial Intelligence
API	Application Programming Interface
BR	Business Requirement
DL	Deep Learning
DoA	Description of Action
IAB	Industrial Advisory Board
JSON	JavaScript Object Notation
ML	Machine Learning
MVP	Minimum Viable Product
TR	Technical Requirement
WP	Work Package
XAI	Explainable Artificial Intelligence