



Industry 5.0 EDGE AI Toolkit - M18

D5.2

Person responsible / Author:	TXT
Deliverable N.:	D5.2
Work Package N.:	WP5
Date:	30/06/2024
Project N.:	101092069
Classification:	Public
File name:	Industry 5.0 EDGE AI Toolkit – M18
Number of pages:	39

The AI REDGIO 5.0 Project (Grant Agreement N. 101092069) owns the copyright of this document (in accordance with the terms described in the Consortium Agreement), which is supplied confidentially and must not be used for any purpose other than that for which it is supplied. It must not be reproduced either wholly or partially, copied or transmitted to any person without the authorization of the Consortium.



Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Health and Digital Executive Agency (HaDEA). Neither the

Status of deliverable

Action	By	Date (dd.mm.yyyy)
Submitted (author(s))	TXT	28.06.2024
Responsible (WP Leader)	SUITE5	28.06.2024
Approved by Peer reviewer	UNP	27.06.2024

Revision History

Date (dd.mm.yyyy)	Revision version	Author	Comments
02.05.2024	v0.1	Gabriele De Luca, Veronica Antonello	ToC
31.05.2024	v0.2	Gabriele De Luca, Veronica Antonello	First Draft with chapter 5 (T5.5)
10.06.2024	v0.3	Nefeli Bountouni, Sotiris Koussouris	Chapter 3 (T5.3) content
18.06.2024	v0.4	Nefeli Bountouni, Sotiris Koussouris	Chapter 4 (T5.4) content
24.06.2024	v0.5	Gabriele De Luca, Veronica Antonello	Final Draft
28.06.2024	V0.6	Gabriele De Luca, Veronica Antonello Nefeli Bountouni, Sotiris Koussouris	Final Version after the review from UNP
28.06.2024	V.1.0	Gabriella Monteleone	Quality Check

Author(s) contact information

Name	Organisation	E-mail	Tel
Gabriele De Luca	TXT	gabriele.deluca@txtgroup.com	
Veronica Antonello	TXT	veronica.antonello@txtgroup.com	
Nefeli Bountouni	SUITE5	nefeli@suite5.eu	
Sotiris Koussouris	SUITE5	sotiris@suite5.eu	
Gabriella Monteleone	POLIMI	gabriella.monteleone@polimi.it	

Table of Contents

1.	EXECUTIVE SUMMARY	6
2.	INTRODUCTION	7
2.1.	SCOPE OF THE DELIVERABLE	7
2.2.	CONTRIBUTIONS TO OTHER WPs AND DELIVERABLES	7
2.3.	STRUCTURE OF THE DOCUMENT.....	7
3.	AI PIPELINE DESIGNER.....	9
3.1.	COMPONENT DESCRIPTION.....	9
3.2.	HIGH LEVEL ARCHITECTURE	9
3.2.1.	<i>AI Pipeline Workbench.....</i>	<i>10</i>
3.2.2.	<i>AI Execution Service.....</i>	<i>11</i>
3.2.3.	<i>Data Manipulation Service.....</i>	<i>11</i>
3.2.4.	<i>Visualization Engine</i>	<i>11</i>
3.2.5.	<i>Visualisation Dashboard.....</i>	<i>12</i>
3.2.6.	<i>AI Models Catalogue</i>	<i>12</i>
3.2.7.	<i>AI Edge Deployment Service.....</i>	<i>12</i>
3.3.	LIST OF AVAILABLE FUNCTIONALITIES.....	13
4.	EDGE AI REFERENCE IMPLEMENTATIONS PORTFOLIO.....	14
4.1.	EDGE AI REFERENCE IMPLEMENTATIONS COLLECTION METHODOLOGY	14
4.2.	PORTFOLIO DESIGN AND CONTENTS	17
4.3.	STRUCTURE OF ASSETS IN THE PORTFOLIO	18
5.	AI-ON-DEMAND PLATFORM ANALYSIS AND PROJECT INTEROPERABILITY	22
5.1.	ANALYSIS OF THE STRUCTURE OF THE NEW PLATFORM RELEASE	22
5.1.1.	<i>AIoD (on Demand) Services</i>	<i>27</i>
5.1.2.	<i>MyLibrary.....</i>	<i>28</i>
5.1.3.	<i>AI Builder</i>	<i>29</i>
5.1.4.	<i>Research Bundles.....</i>	<i>31</i>
5.1.5.	<i>Reproducibility.....</i>	<i>31</i>
5.1.6.	<i>RAIL.....</i>	<i>31</i>
5.1.7.	<i>Integrate: Service Builders.....</i>	<i>33</i>
5.1.8.	<i>AIoD metadata Catalogue.....</i>	<i>34</i>
5.1.9.	<i>AIoD open API.....</i>	<i>34</i>
5.1.10.	<i>Integrate with AIoD.....</i>	<i>35</i>
5.2.	INTEROPERABILITY BETWEEN THE AI REDGIO PLATFORM AND THE AIoD PORTAL.....	35
6.	CONCLUSION.....	39

Figures

Figure 1 – AI Pipeline Designer High-Level Architecture	10
Figure 2 – AI Pipeline Workbench Interface – Graph View	11
Figure 3 – Edge Deployment Guidelines.....	12
Figure 4 – Reference Implementations Collection Phases	15
Figure 5 – AI REDGIO 5.0 Portfolio Asset Taxonomy	15
Figure 6 – AI REDGIO 5.0 Portfolio – Asset Information Collection Template	16
Figure 7 – Process of Asset Collection and Onboarding in Portfolio	16
Figure 8 – Main Menu of AI REDGIO Portfolio Wiki	17
Figure 9 – AI REDGIO 5.0 Portfolio – Asset Template.....	19
Figure 10 – AI REDGIO 5.0 Portfolio – Asset Example	21
Figure 11 – Architecture of the latest AI-on-Demand platform release	23
Figure 12 - REST API and Python SDK: integration options for enhanced accessibility	24
Figure 13 - Development roadmap of AloD platform: implemented features and future releases	24
Figure 14 – Interaction with the platform through REST API and SDK	26
Figure 15 - Discover the available assets in the AloD Assets Catalogue	29
Figure 16 - AloD Marketplace: explore the available assets	30
Figure 17 – My Models: a collection of the user's uploaded models.....	30
Figure 18 - Process that takes place when an experiment template is created	32
Figure 19 – Experiment creation process	32
Figure 20 – AI REDGIO portal integration with AloD Metadata Catalogue.....	35
Figure 21 – AloD available assets Information Model.....	36
Figure 22 – AI Asset base attributes	38

Tables

Table 1 - AI Assets Attributes.....	36
-------------------------------------	----

Abbreviations and Acronyms:	
AI	Artificial Intelligence
AIoD	AI-on-Demand
oD	On Demand
API	Application Programming Interface
REST	REpresentational State Transfer
SDK	Software Development Kit
SMEs	Small and Medium Enterprises
UI	User Interface
URL	Uniform Resource Locator
UX	User Experience
ML	Machine Learning
PC	Personal Computer
VR	Virtual Reality
DB	DataBase
CWL	Common Workflow Language
HTTP	Hypertext Transfer Protocol

1. Executive summary

Deliverable D5.2 presents the results, achieved at M18, regarding three different tasks of the project: T5.3 - “Cloud-to-Edge AI Pipeline Lifecycle Management Platform for Industry 5.0”-, T5.4 - “Edge AI Reference Implementations for Industry 5.0”- and T5.5 - “Interoperability with AI-on-demand platform”-. Due to the OTHER nature of the deliverable, D5.2 is an accompanying document of the artifacts designed and implemented in the context of the three tasks above mentioned.

The content of D5.2 includes a comprehensive analysis of a collaborative intelligence platform that enables users to enhance their skills and abilities in using AI assets, focusing in particular on: interoperability with the AI-on-demand platform at different levels, a cloud-based tool to manage, design, deploy and execute AI pipelines, and a library of Edge AI reference implementations.

First, the AI Pipeline Designer platform is analysed, an interface designed to meet the needs of different users and to facilitate the development, training, execution and deployment of AI models. The description includes an analysis of its components, followed by a detailed overview of its architecture and available services. The platform also provides a catalogue of AI models and a comprehensive list of all available functionalities.

This is followed by an analysis of the AI REDGIO 5.0 Edge AI Reference Implementations for Industry 5.0 portfolio, which collects, organises and publishes Edge AI assets for the manufacturing domain. First, the methodology for collecting Edge AI reference implementations is addressed, then the portfolio design and contents are illustrated, and finally the structure of the assets in the portfolio is reported.

In the final part, the analysis focuses on the newly released version of the AI-on-demand platform. In particular, the structure of the platform is examined, with an emphasis on the technical tools that are most valuable from a technical perspective. Special attention is given to the interoperability of the project.

The analysis of the platform structure will be the starting point for the implementation of the interoperability with the AI-on-demand platform, which will be the main topic of the task T5.5. In particular, the task will focus on ensuring the exposure and interoperability of the AI REDGIO platform with the AI-on-Demand platform. Through this integration, assets generated as a result of the work carried out for the project will be added to the existing resources available on the platform. This contribution will increase the overall value of the platform, providing users with a wider range of tools to support AI development and deployment, exploiting different AI assets created to satisfy specific manufacturing needs.

2. Introduction

2.1. Scope of the Deliverable

This deliverable reports the status of the implementation of the Cloud-to-Edge AI Pipeline Lifecycle Management Platform, accompanied by the Edge AI reference library for selected manufacturing problems, and the interoperability with the AI-on-Demand platform, at M18.

Through the AI Pipeline Designer platform, users have the possibility of design, train, execute and deploy multiple AI models or pipelines to create specific solutions related to a specific objective. This platform provides a user-friendly interface that guides the user through the implementation of complete pipelines for specific problems. Doing so it facilitates the smooth onboarding, creation and deployment of AI models for several categories of users.

The AI REDGIO 5.0 Portfolio integrates solutions and assets from the expertise of the AI REDGIO 5.0 partners, specific knowledge developed within the AI REDGIO 5.0 project, and existing open-source solutions outside the project. The portfolio is intended both for AI REDGIO partners, who can benefit from the solutions created by other partners for their WP6 activities, and for external users, providing a valuable resource for any external stakeholder wishing to implement their own Edge AI projects.

Better accessibility to datasets and artificial intelligence models from various platforms has been achieved thanks to the recently released version of the AI-on-Demand (AloD) platform. This release is related to the fact that the requirements changed from just creating a catalogue of different AI resources to build a platform that users can integrate with. The different types of assets available on the platform can be used to overcome some of the challenges of using AI, making the AI tools more accessible.

The strategic integration of the platform with the AI REDGIO portal can improve the usefulness and effectiveness of the platform by leveraging the AI tools explored in several processes, providing users with additional tools.

The in-depth analysis of the platform structure and the available technical tools are crucial for the subsequent analysis of the integration of the AloD platform with the AI REDGIO portal and the analysis of the integration of the Pipeline Designer with the AI Builder, which will be the focus of the next iteration of the project.

2.2. Contributions to other WPs and deliverables

The results of this deliverable have been influenced by the requirements gathered in WP2 and WP6. In fact, the AI requirements identified following the in-depth analysis of the experiments have facilitated the definition and enhancement of the project technology for implementing AI at the edge.

Furthermore, the technical assets provided in WP4 and WP5 are supported by the AI toolkit, which allows the integration of existing functionalities with AI.

Finally, the results described are currently undergoing validation and verification analysis in T5.6, with the associated results being documented in D5.3.

2.3. Structure of the Document

This document is divided into three different sections. Each section contains the analysis of a tool related to one of the specific tasks described above.

The first section of this document is focused on the AI-on-Demand Pipeline Designer, the interface through which AI models can be designed, tested, trained, executed and deployed. This tool fulfils the objective of the T5.3: to design and deploy a pipeline management platform that enables to design, schedule and monitor different artificial intelligence pipelines in the cloud and orchestrate their execution at the edge or in a hybrid cloud-edge context, using the open hardware platform delivered in T5.2. Thus, this task involves federated learning capabilities for a set of fundamental machine learning and deep learning algorithms to address multiple problems.

The platform allows the integration of already trained models and pipelines to address specific manufacturing needs and is delivered as an Edge AI reference implementation on T5.4. The first part of the document provides a general description of the tool, as well as an analysis of its architecture and functionalities. It also describes how to use it and the various ways in which this tool can facilitate and improve the creation of AI pipelines.

The second section of this document relates to the portfolio of Edge AI reference implementations. This tool fulfils the objectives of T5.4, which aims to collect reference implementations of Edge AI models and assets for manufacturing business cases, drawing on open-source initiatives and partner expertise. These reference implementations are enhanced by specific Edge AI models and pipelines developed within the AI REDGIO 5.0 project to address the concrete problems identified in the experiments. In particular, this section of the document describes the functionality of the portfolio and the purpose behind its implementation. The process of monitoring and gathering assets is also illustrated. In addition, the portfolio designs and their context, as well as the structure of the assets within the portfolio, are analysed to provide a more comprehensive understanding.

During the first phase of the project, the AI-on-Demand platform underwent a profound change in its philosophy, with two different versions of the platform available at the time of writing this document. As the development process of the latest version is still ongoing, it was necessary to analyse the recently released version of the AI-on-Demand (AIoD) platform in detail, particularly focusing on the most relevant available technical tools. In addition, some of the features that characterise the platform, such as physical resources and AI tools, are not implemented yet as the development of the platform is based on later releases.

The final section of this document details the analysis carried out to ensure the compliance of the project assets with the AI-on-Demand platform and the preliminary results obtained. This part of the document reports the interoperability analysis between the AI-on-Demand platform that will be the central core of T5.5, which focuses in particular on the interoperability with the AI REDGIO portal, aiming to facilitate the addition of project-derived assets to the existing resources available on the platform, thereby providing users with an extended range of assets.

3. AI Pipeline Designer

3.1. Component Description

Developed under T5.3, the AI Pipeline Designer platform is the interface through which the experiments are able to design, experiment with, train, execute and deploy AI models that are integral parts of the use cases they are building.

The AI Pipeline Designer platform addresses the needs of different types of users (i.e. data scientists, technical users, business users) in terms of AI and ML execution, from executing simple data manipulation functions (e.g. filters and aggregations) and applying Artificial Intelligence models for the manufacturing domain analytics, to creating visualisations and reports to highlight insights extracted from datasets and from analytics processes, as well as exporting the outputs of these analyses through interfaces that can be consumed by other systems. Moreover, the platform also allows the deployment of certain models on Edge devices (depending on the complexity of the pipeline and the resources required to effectively execute a model), while it is being prepared to ingest models from the AIoD platform, as well as to export models back to the latter.

The platform provides an user-friendly interface that guides the user through the creation of complete pipelines, from selecting and configuring the input data, to the step-by-step processing and exporting the results in the desired form. Data manipulation and analysis functions are provided in the form of easily configurable blocks that can be combined to obtain a data analysis pipeline. The different algorithms that the user can select range from the mainstream algorithms offered by popular frameworks to customised, pre-trained algorithms for specific problems (for example algorithms that can be imported from other systems, such as the AIoD platform, assuming they are compatible)

As identified above, the execution of the a data analysis pipelines can be configured for where (cloud or edge) and when it will be planned to be executed, allowing the platform to be used either a) for supporting the cloud-based execution of models (for experimentation as well as for model training purposed but also for the execution of resource-heavy analytics or for cases that do not require the edge execution of a model), and b) for deploying specific models to Edge devices in order to facilitate the smooth onboarding of AI models to such devices for the cases where edge AI execution makes sense.

3.2. High Level Architecture

In more detail, the AI Pipeline Designer Platform is composed, as can be seen in the architecture illustrated in [Figure 1](#), by the sub-components that are described above, which are all orchestrated using a Kubernetes Engine.

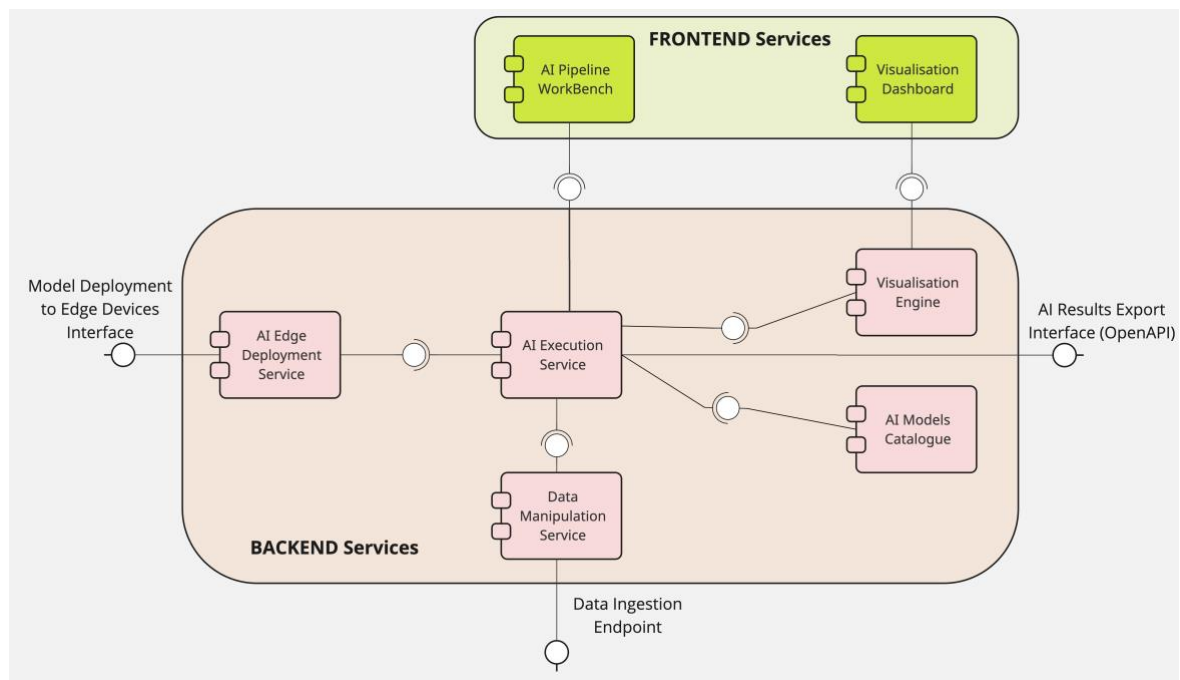


Figure 1 – AI Pipeline Designer High-Level Architecture

3.2.1. AI Pipeline Workbench

The AI Pipeline Workbench is responsible for the creation of AI pipelines at “design” time. The creation of a data and AI analytics pipeline includes the following steps: (a) selection of the data, (b) configuration of data manipulation functions using the provided user interface, (c) training, application and evaluation of models, (d) the selection of the output data, (e) specification on the actions to be performed with the output data (such as storage, retrieval through APIs and visualization). In addition, the overall execution configuration (e.g. schedule, AI framework, re-training options) is considered as part of the design process of the data analytics pipeline. The Designer is the main UI that is provided to the user. A screenshot of it is provided in the following figure.

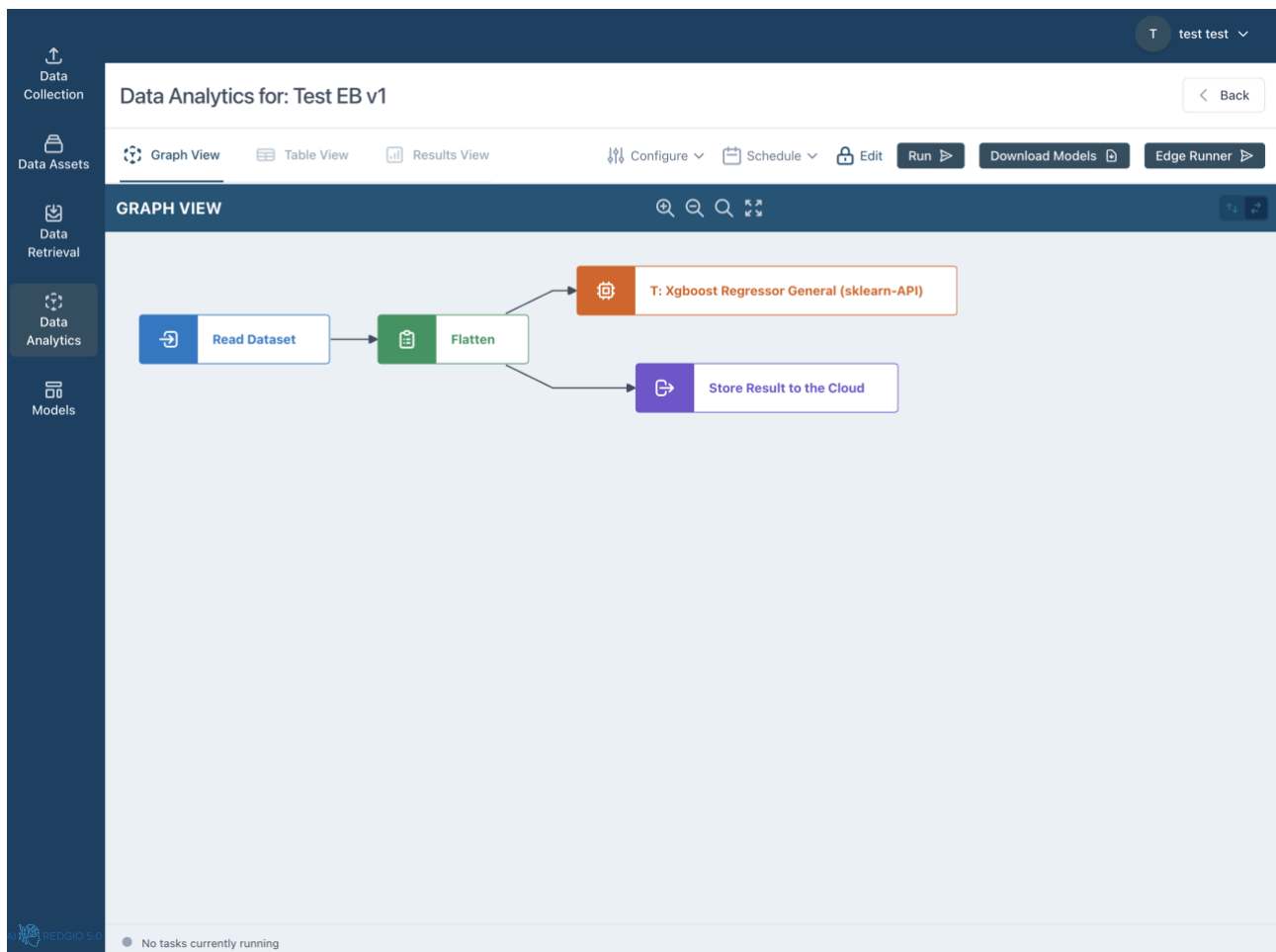


Figure 2 – AI Pipeline Workbench Interface – Graph View

3.2.2. AI Execution Service

The AI Execution Service is responsible for executing different AI pipelines created in the AI Pipeline Workbench. It achieves this by invoking specific functions based on the configuration parameters in the chosen AI framework/library.

3.2.3. Data Manipulation Service

The Data Manipulation Service provides data manipulation functionalities, including filters, conditional column creations, aggregations, merging with other datasets, performing mathematical computations, removing null values, etc. These are used to transform a dataset in the appropriate format to serve the needs of users that want to preprocess their data prior to providing them as input to the AI algorithms.

3.2.4. Visualization Engine

The Visualization Engine is the component that offers functionalities to allow the end users: (a) to create custom visualizations of the data/results derived from the data ingestion/analytics pipeline they want to dive into (by configuring various built-in charts); (b) to consolidate and manage multiple charts in a user-friendly

manner through functions for their dynamic export, etc. This component is a backend service that takes instructions from the Visualisation Dashboard, fetches the data from the executed AI pipelines, transform them into the appropriate structures to be consumed by the Dashboard and then serving those back to the dashboard elements to be visualised

3.2.5. Visualisation Dashboard

The Visualisation Dashboard is the component that consumes the outputs coming from the visualisation engine and is responsible for constructing the different graphs and visualisations that are presenting the outputs of the AI models. This component offers the user the ability to choose from several graphs and visualisation templates and after selection enables them to configure those to display the information that is the output of the AI Engine.

3.2.6. AI Models Catalogue

The AI Models Catalogue is the component responsible to handle the collection or inventory of various mainstream AI models defined within the context of the AI REDGIO 5.0 project by the AI Pipeline Designer Platform, facilitating analytics using structured data sources. This catalogue serves as a repository of out-of-the box algorithms implementations that can be utilized for different purposes (e.g. classification, clustering, regression, etc) in order to perform specific functions based on experiments' analytics needs. Other AI models, which are to be provided as open source and are specific to the experiments (for example computer vision algorithms, etc), will be provided in the project's wiki, and made available publicly. The users of the AI Analytics platform are able to browse the catalogue to select and incorporate the most suitable AI models into their configuration (as AI blocks), saving time and resources by leveraging existing expertise and knowledge.

3.2.7. AI Edge Deployment Service

The AI Edge Deployment service is responsible for the packaging of AI models and proceeding with their deployment to edge devices that have been registered in the platform, so they can receive this code. The service initially evaluates the feasibility of the AI model to be executed over an edge device (for example checking the ability of the analytics framework to be executed at the edge device). Once this evaluation is over, in case the service gets the green light, it proceeds with packaging the model and generating a docker container, that can be directly downloaded by the edge device and triggered then for execution by the platform. This information is provided by the platform as seen in the figure below.



Figure 3 – Edge Deployment Guidelines

3.3. List of Available Functionalities

The list of functionalities offered by the AI Pipeline Designer in the first version provided under deliverable D5.2 are provided below. These features will be documented as user stories in D5.3.

- AIPD_1 – Definition of re-usable and customisable data treatment blocks: The AI Pipeline Designer wraps the functions provided by the Data Manipulation Service in configurable and pipeline blocks, allowing the user to select the block that offers the desired functionality, parameterise it and combine it with other blocks to generate a pipeline
- AIPD_2 – Definition of re-usable and customisable analytics models blocks based on mainstream algorithms: For users that want to have more flexibility in implementing their own models, the workbench offers some generic algorithms that can be trained and configured in a more flexible way, but with much more effort to be spent on experimentation.
- AIPD_3 – Re-usable and customisable data input blocks: To allow the users to use the aforementioned blocks, appropriate input blocks are also provided. The input blocks indicate data assets that an organisation has at its disposal.
- AIPD_4 – Re-usable and customisable data output blocks: The AI Pipeline Designer allows the user to define output blocks according to their own needs in order to visualize the results through custom diagrams that are created and saved with the help of the Visualization Engine, to download the results as a file or to retrieve the results via Open APIs
- AIPD_5 – Configuration and validation of AI pipelines: After creating and connecting a series of blocks in a pipeline, the user needs to provide the execution details of the defined pipeline. The execution of a pipeline can be practically scheduled to run once, at specific intervals or upon a triggering event in different locations.
- AIPD_6 – Deployment of AI models on Edge devices: The user is able to select deployment at edge devices and is able to trigger the execution of those models.
- AIPD_7 – Update of data analysis pipelines: After a data analysis pipeline is finalised, the user can revise a part of the configuration in order to leave some flexibility for updates, but without compromising the consistency of the results.
- AIPD_8 – Insights into the execution of AI pipelines: Every time the Data Manipulation Service and the AI Execution Service are executed, they collect certain metrics about the outcome and different performance aspects of the data analysis pipeline that are displayed in the Analytics Workbench. The data asset consumer may view the report for the latest run, but also for all successful runs.
- AIPD_9 – Registration of externally trained compatible AI models: The AI Pipeline Designer allows users that have trained ML/DL models outside the AI REDGIO 5.0 Platform yet in compatible libraries, to register them in the platform.

4. Edge AI reference implementations portfolio

The AI REDGIO 5.0 Edge AI Reference Implementations for Industry 5.0 Portfolio (hereinafter called ‘AI REDGIO 5.0 Portfolio’) collects, organises and publishes Edge AI Reference Implementations (i.e. assets) for the manufacturing domain. It combines solutions and assets coming from the AI REDGIO 5.0 partners’ background, existing open-source solutions outside the AI REDGIO 5.0 project, and foreground knowledge (e.g. specific models, pipelines and datasets) created as part of the AI REDGIO 5.0 project.

The AI REDGIO 5.0 Portfolio is targeted both towards the AI REDGIO partners (experiments, technical partners and researchers), that could benefit from solutions developed by other partners for their WP6 activities, while it is also at the disposal of the general public, aiming to constitute a useful resource for any external stakeholder in implementing their own Edge AI projects.

For this reason, it has been deployed as an online wiki and is comprised of carefully selected assets relevant to Edge AI for Industry 5.0, exposing only information that can become public, while in its second release [M30] it will be enhanced with explorable User Journeys, for improved user experience, going beyond asset cataloguing.

The deployed AI REDGIO 5.0 Portfolio is available through the link: https://wiki.ai-redgio50.s5labs.eu/index.php?title=Main_Page.

4.1. Edge AI Reference Implementations Collection Methodology

The AI REDGIO 5.0 Portfolio has a twofold purpose: (a) assisting users in their Industry 5.0 endeavours through the provision of multi-faceted Edge AI assets that cover a wide span of AI types and manufacturing problems and (b) collecting and providing visibility to the tangible AI-related outcomes of the AI REDGIO 5.0 project.

Thus, the asset collection and selection methodology has been carefully designed to allow continuous monitoring and identification of relevant outcomes emerging from the activities of the experiments (WP6) and the technical developments (WP4 and WP5).

Following the project’s activities, anticipated development outcomes and the maturity of experiments, the monitoring and collection of assets can be summarised in three phases:

- Phase I: Early collection and population of the AI REDGIO 5.0 Portfolio with assets coming from the background of the AI REDGIO 5.0 partners (e.g. other EU or internal projects, or open-source solutions).
- Phase II: Enhancement of AI REDGIO 5.0 portfolio with initial results from the AI REDGIO 5.0 technical activities (e.g. work under WP4 and WP5) and models or technical developments in the context of the first phases of the SME and DF experiments (WP6)
- Phase III: Extension of the AI REDGIO 5.0 portfolio with assets from the second release of the AI REDGIO 5.0 technical work packages and experiments activities (WP6).

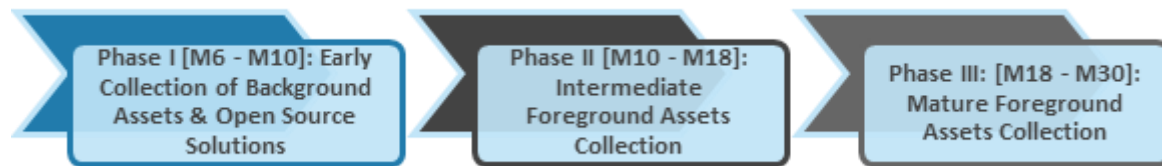


Figure 4 – Reference Implementations Collection Phases

An initial taxonomy of AI assets has been constructed to provide the underlying framework for the asset categorisation in the AI REDGIO 5.0 Portfolio. This taxonomy was structured based on a combination of taxonomies proposed in the industry and literature with a focus on the manufacturing and Industry 5.0 domain (e.g. categorisation of assets in the AI-on-Demand platform¹), and the Edge AI technologies in the scope of AI REDGIO 5.0 (e.g. from early interviews conducted with experiments for the identification of their AI needs in the context of WP5 and reported in deliverable D5.1 – ‘Industry 5.0 EDGE AI Foundations’).

The AI REDGIO 5.0 Portfolio taxonomy has already been used in the deployed wiki for the categorisation of published assets and will be updated as needed during the course of the project in order to cover new aspects emerging from the collected assets.

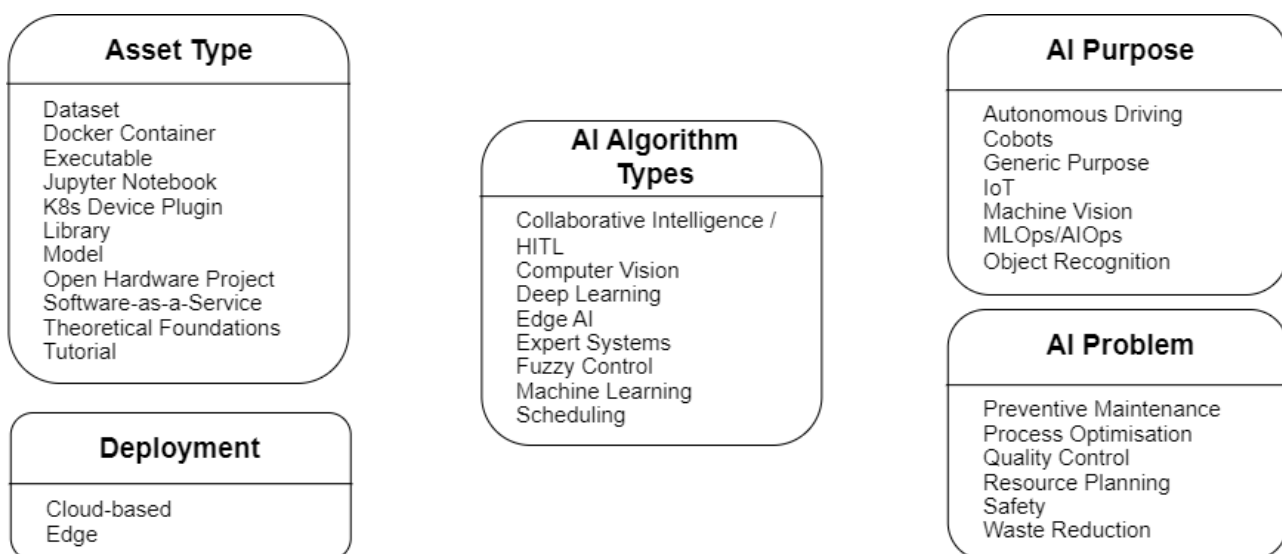


Figure 5 – AI REDGIO 5.0 Portfolio Asset Taxonomy

Apart from the descriptive and technical information related to the assets, additional information should accompany them in order to facilitate their usage by interested users exploring the portfolio. This includes among other: licence information, usage guides and instructions, images and any other material illustrating the solution, acknowledgement (since the asset might originate from other projects from the background of partners), contact information. All this information has been captured in the asset information collection template, that is used to support the asset collection process.

¹ <https://www.ai4europe.eu/research/ai-catalog>

The template has been designed to be applicable to diverse asset types (e.g. executables, models, datasets, etc), and adaptable, while keeping a unified structure that can be easily transported to the AI REDGIO 5.0 Portfolio wiki. It consists of the sections reported in [Figure 6](#).

- **Asset's Short Description:** <brief explanation of the purpose of the specific reference implementation>
- **Technical Information:** <brief presentation of the asset from a technical perspective. Depending on the asset type, this could include on or more of the following (indicative, non-exhaustive list): the asset specifications, the features / functionalities, a high-level architecture, underlying/related technologies (e.g.: sk-learn, Tensorflow, Spark MLlib, etc.) etc.>
- **Usage Aspects:** <a guide for the stakeholder that wants to use the asset. The section could include: usage walkthrough / instructions, links to instructions/demonstrations/other resources available online>
- **Relevant AI REDGIO 5.0 Experiment:** <if applicable to the asset: the name of the Experiment that (will) use / demonstrate the asset in their experiment>
- **Maturity:** <the applicable level of maturity for the specific asset: Proof-of-Concept, Ongoing Development, Already Implemented/Available, Other (please specify)>
- **Licence:** <the applicable asset licence: Proprietary, Restricted, Open source. The specific Licence (e.g. Apache License Version 2.0) should also be specified>
- **External Resources:** <Online resource where the asset is available (e.g. GitHub repo)>, <Creator>, <Contact information> (optional)
- **Acknowledgement:** <if the asset was created in the context of another project / initiative, a relevant statement should be included here>
- **Categories:** <categories that are relevant to the asset. Options available based on the defined taxonomy. 'Other – please specify' option is also available and defined taxonomy can be extended based on proposed additional categories (after review)>
- **Images:** < screenshots, images relevant to the asset, with the relevant caption/description>

Figure 6 – AI REDGIO 5.0 Portfolio – Asset Information Collection Template

The asset collection process is initiated periodically during each phase and is supported by the information collection template that is circulated to the AI REDGIO 5.0 partners. It comprises the steps illustrated in the following figure ([Figure 7](#)).



Figure 7 – Process of Asset Collection and Onboarding in Portfolio

After each asset collection round is completed, it is followed by an asset selection and refinement phase. The collected assets are reviewed and assessed based on their relevance to the scope of the AI REDGIO 5.0 Portfolio and added-value for prospective users of the Portfolio. Afterwards, the selected assets are onboarded on the AI REDGIO 5.0 Portfolio wiki and the partners that provided the assets are invited to review the relevant entries and provide any comments for adaptations.

4.2. Portfolio Design and Contents

The AI REDGIO 5.0 Portfolio wiki has been designed to facilitate asset exploration and identification of assets of interest for end-users. This is achieved through the clear content structure, asset categorization and categories exploration (e.g. through categories tree), text-based search.

The content published through the first version of the AI REDGIO 5.0 Portfolio wiki, available in [M18] is structured in four pages (Figure 8).

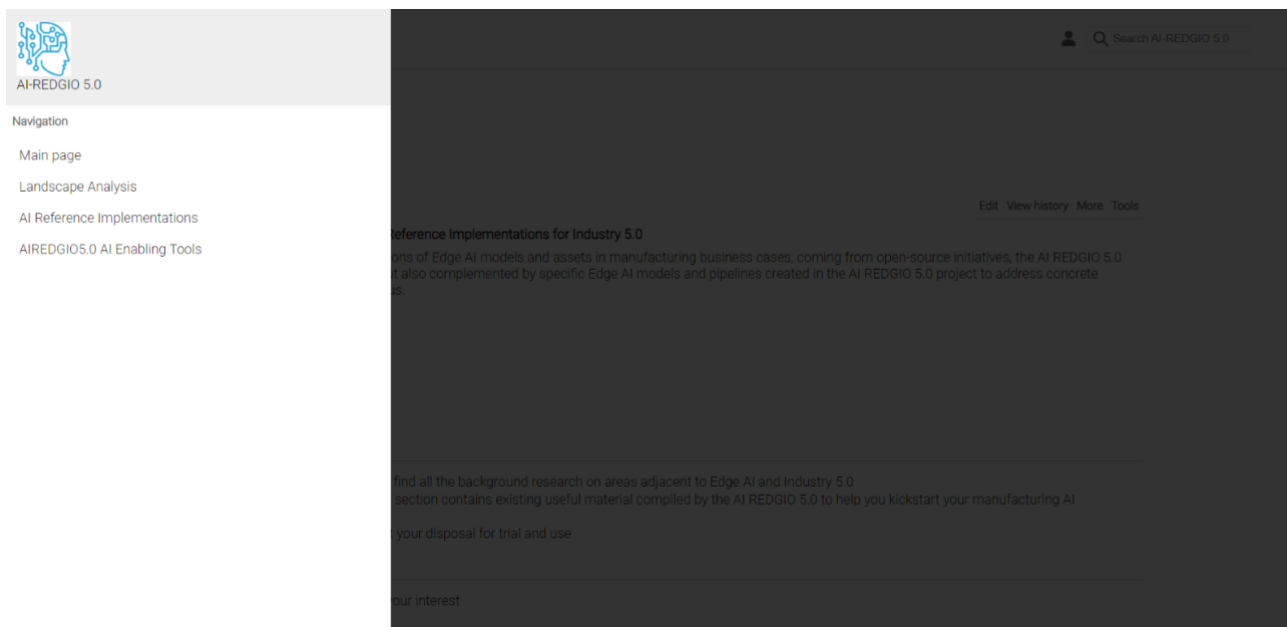


Figure 8 – Main Menu of AI REDGIO Portfolio Wiki

- **Main Page:** The home page of the portfolio. It hosts introductory information for the user, including (a) the scope of the AI REDGIO 5.0 Portfolio, (b) a brief outline of the content of the other items of the main menu and quick links to directly navigate there (*'Getting Started'* section), (c) the overview of the portfolio taxonomy (section 4.1) through a categories' tree that contains quick links to the published assets grouped per category (*'Asset Categories'* section), (d) the five assets / pages that were last updated/added in the wiki (*'Latest'* section), (e) acknowledgement to the AI REDGIO 5.0 project and link to the official project website (*'About AIREDGIO 5.0'* section).
- **Landscape Analysis:** Contains theoretical foundation assets that encapsulate findings from research on fields adjacent to the scope of the AI REDGIO 5.0 Portfolio (e.g. Collaborative Intelligence, MLOps and AIOps, Open Hardware solutions)
- **AI Reference Implementations:** The compilation of background and foreground AI assets that can help the users kickstart their Industry 5.0 journey. Indicative contents of this page include assets

developed in the context of WP6 experiments (e.g. models, datasets, test infrastructures), background assets brought by the AI REDGIO 5.0 partners (e.g. plugins from internal projects, dashboards created in other EU-funded projects)

- **AI REDGIO 5.0 AI Enabling Tools:** A page hosting resources relevant to the AI solutions developed under the technical work packages of AI REDGIO 5.0 (e.g. the Collaborative Intelligence platform).

4.3. Structure of Assets in the Portfolio

The published assets in the portfolio follow a similar structure that is adapted as needed to fit the specificities of each asset type and the information provided. An Asset Template page has been created

that is instantiated with the information provided for new assets onboarded to the Portfolio (Figure 9).


AI-REDGIO 5.0



Asset Template

From AI-REDGIO 5.0

Page
Edit View history More Tools

A page that serves as a basic template for the asset pages

Contents [hide]

- 1 Asset Description
- 2 Features
- 3 Usage
 - 3.1 Walkthroughs
 - 3.2 Licence
- 4 External Resources
- 4.1 Acknowledgement
- 5 References
- 6 Relevant Categories



image 1: caption

Asset Description

 [edit]

Lorem ipsum dolor sit amet, consectetur adipiscing elit ^[1], sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.



Image 2: Screenshot from the Collaborative Intelligence Platform

Features

 [edit]

Feature	Description
Example	Example
Example	Example
Example	Example

Usage

 [edit]

How to deploy and use this asset:

- Step 1
- Step 2
- Step 3
- Step 4

The asset is Available at [Kaggle](#).

Walkthroughs

 [edit]

- Instructions available [here](#)
- Detailed step-by-step demonstration can be found [here](#)

Licence

 [edit]

MIT License, licence link [here](#)

External Resources

 [edit]

- Created by [Suite5 Data Intelligence Solutions](#)
- Contact [\[contact@suite5.eu email\]](#)

Acknowledgement

 [edit]

This work is an was originally created in the context of xxx Project

References

 [edit]

1. T. Ivanov, D. (2023). The Industry 5.0 framework: Viability-based integration of the resilience, sustainability, and human-centricity perspectives. International Journal of Production Research, 61(5), 1683-1695.

Relevant Categories

 [edit]

Category: Xxxx

Figure 9 – AI REDGIO 5.0 Portfolio – Asset Template

The Asset Template page consists of the following elements and sections that follow the structure and contents of the information collection template (Section 4.1):

- **Asset title:** the name of the asset
- **Contents:** the list and scrollable links to the underlying sections in the asset page. Sections are identified based on used Headings in the specific wiki page

- **Asset Description:** brief description of the asset's purpose and relevance to AI for manufacturing
- **Features:** an outline of the asset offerings (e.g. its functionalities)
- **Usage:** contains the usage-related information, including walkthroughs and instructions, demo material, and license information
- **External Resources:** contains the links to external repositories or other websites hosting the asset (e.g. a GitHub repository), contact information to the partner that provided the asset
- **Acknowledgement:** any required acknowledgement and relevant links (e.g. reference to another project where the asset was originally created)
- **References:** any citations and relevant resources
- **Relevant Categories:** the categories from the AI REDGIO 5.0 Portfolio taxonomy that are applicable to this asset

The asset template is adapted according to the specificities of its asset. The following example from the 'Quality Control in Industry with CV and TinyML' asset ([Figure 10](#)) illustrates how the template has been instantiated and populated with the content provided by the relevant AI REDGIO 5.0 partner.

Quality Control in Industry with CV and TinyML

From AI-REDGIO 5.0

Page

Edit View history More Tools

Quality control on production lines with computer vision and TinyML involves using computer vision algorithms and TinyML technology to automatically inspect and detect defects in products as they move along the production line

Contents [hide]

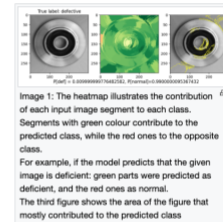
- 1 Asset Description
- 2 Features
- 3 Usage
 - 3.1 Walkthroughs
 - 3.2 Licence
- 4 Resources
 - 4.1 Acknowledgement
- 5 Relevant Categories

Asset Description [edit]

Computer Vision (CV) can be used to analyze images of products and compare them to a "good" or "reference" image to identify any defects. These algorithms can be trained to detect a wide range of defects, including scratches, dents, misalignments, and missing components. Additionally, Computer Vision algorithms are able to work in real-time, allowing it to detect defects in products as they move along the production line and flag them for further inspection or rejection.

TinyML is a field that involves developing machine learning models that can run on small, resource-constrained devices such as microcontrollers. In the context of quality control on production lines, TinyML can be used to enable the **CV algorithms to run on embedded devices**, such as cameras or sensors, that are integrated into the production line. This allows the system to process images and make decisions about defects without the need to send data to a separate computer for analysis.

By combining computer vision algorithms and TinyML technology, it's possible to create a **real-time, automated quality control system** that can detect defects in products as they are produced, improving the overall quality and efficiency of the production process.



Features [edit]

This asset includes a Jupyter notebook that presents a complete pipeline for:

Feature	Description
Exploratory Data Analytics (EDA) on image data	Pandas Dataframe, exploring and visualize the datasets and its metrics.
Data preparation and augmentation	Scale the values down to a smaller range, usually between 0 and 1. We use the <code>tf.keras.layers.Rescaling</code> layer in TensorFlow to accomplish this standardization and adjust the channel values of each image to be in the [0, 1] range. Data augmentation increases the diversity of the training set by applying random (but realistic) transformations, such as image rotation.
Image classification with TensorFlow	with Deep learning (CNN) models
Models evaluation	Confusion Matrix
Model interpretation predictions	with LIME
Transformation for use in embedded devices	transformation to TFLite format
Post-training quantization	Post-training quantization is a static quantization technique where the quantization is applied to the already trained model. This method can be applied to both weights and activations, or just the weights.
Quantization aware training	Quantization aware training (QAT) involves training a model with knowledge of the quantization process that will be used to reduce the precision of the model's weights and activations during deployment.
Overall evaluation	Evaluation based on metrics such as Accuracy, Recall, Precision, F1-Score, Inference Time and Size.
Real-time usage of the developed models	with a python script <code>opencv_object_tracking.py</code>

Usage [edit]

This Notebook assumes that there is available a dataset of labeled images of a product and demonstrates how the user can perform the following:

- Dataset overview
- Data preparation and augmentation
- CV model creation for image classification
- Model transformation to TinyML using TFLite
- Comparison between the original and TFLite mode.

The asset is Available (Note: For demonstration purposes, we use the casting product image data for quality inspection dataset available at [Kaggle](#). However, a similar logic could be applied to other industries and product lines.)

Walkthroughs [edit]

- Instructions available [here](#)
- Detailed step-by-step demonstration of the Quality Control asset can be found [here](#)

Licence [edit]

MIT License, licence link [here](#)

Resources [edit]

- Notebook available in [Github](#)
- Created by [ExpertAI-Lux S.à r.l](#)
- Contact george.fatouros@expertai-lux.com / george.makridis@expertai-lux.com

Figure 10 – AI REDGIO 5.0 Portfolio – Asset Example

5. AI-on-demand Platform Analysis and Project Interoperability

5.1. Analysis of the structure of the new platform release

The new version of the AI-on-Demand (AloD) platform² was recently released, last year, in 2023. It brought better accessibility to datasets and artificial intelligence models from multiple platforms, provided an SDK (Software Development Kit) and an open API (Application Programming Interface) to build services on top of it, and improved some functionality of existing services, such as the AI Builder (formerly AI4Experiments). This release is related to the fact that the requirements changed from just creating a catalogue of different AI resources (models, datasets, papers and so on) to build a platform that users can integrate with, therefore, within the EF Europe project was created a new concept.

The aim of creating the platform is to provide a tool that can help researchers and practitioners in SMEs (Small and Medium Enterprises) and other industries to exploit AI in many ways. The platform can be used as a real platform to build services on top of and can also be used to obtain different kinds of AI assets, datasets, models and so on.

The different types of assets available on the platform can be useful in overcoming some of the challenges of using AI. First, working with AI often requires large amounts of data, but obtaining reliable, high-quality data can often be a challenge. In addition, the platform aims to provide massive computing power, as AI is a powerful tool, but it also requires powerful hardware to train. Moreover, if several parties make their resources available, it is crucial to ensure that these resources are only provided to the people selected by the provider. Hence, there is the need to have a reliable identity provider in the platform.

The AI-on-Demand (AloD) platform is a community-driven channel designed to enhance European research and innovation in the field of artificial intelligence (AI), while guaranteeing the European brand of quality, reliability and explainability. Open and easily accessible, AloD facilitates knowledge sharing, research experimentation and the development of advanced solutions and technologies in the field of artificial intelligence. The AloD platform enables the AI community to explore the potential and the opportunities related to AI applications, to share knowledge, services or tools related to AI, and to take advantage of the many available resources. The platform aims to accelerate the adoption of AI-based solutions by facilitating access to AI, providing tools that facilitate its understanding. Moreover, by enabling users to be part of the platform, it allows them to contribute to its growth and evolution by incorporating their needs into new developments.

Users can benefit in several ways from the platform:

- (i) developing or technically implementing different components, services or tools
- (ii) using it as a dissemination channel to share research results
- (iii) accessing courses and educational resources on AI
- (iv) accessing relevant information, resources, datasets, tools or services
- (v) collaborating with other users.

In particular, the users that can benefit from the platform can be categorized in two different types:

- The consumers, that can consume the AI assets and the services that are available on the platform.

² <https://aiod.eu/>

- The prosumers, or producers of services, that can use, for instance, platforms API to build their services in an easier way, or they can even provide the services directly through the platform and potentially reach a larger audience.

The platform aims to be decentralized, which means that there will not be one central point that will represent the platform, and synchronized, so potentially there will be many different deployments of the platform and these nodes will be aimed to collaborate to share the metadata and the resources. Such an architecture provides, to those who deploy on their own, the full control over the node itself. The following figure (Figure) shows a representation of the structure of the recently released version of the AI-on-Demand platform. The root nodes connect different leaf nodes, that, instead, are related to one root node only.

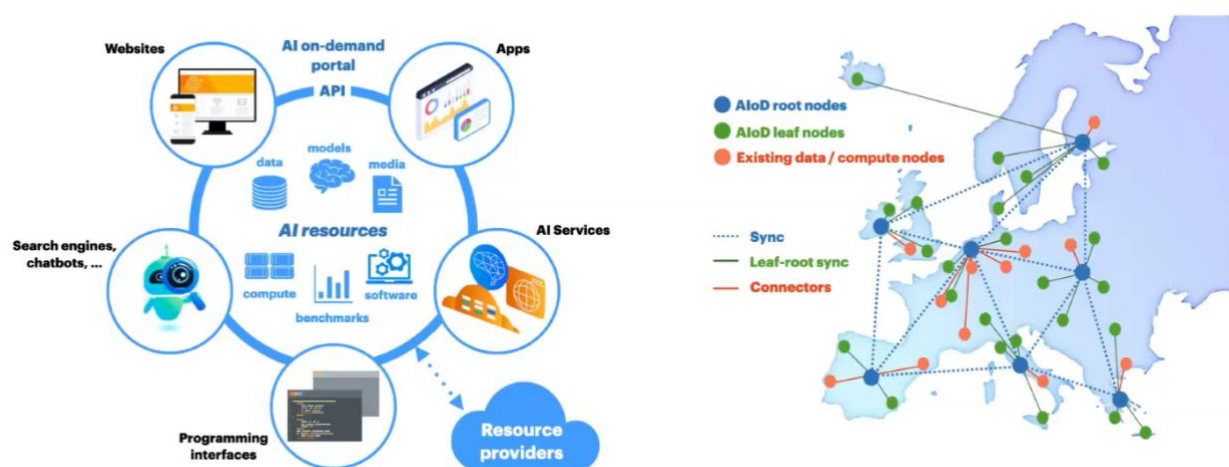


Figure 11 Architecture of the latest AI-on-Demand platform release

The services that are built on top of the platform will be included in the AI-on-Demand website, which is the entry point to the platform itself and allows to directly access the needed service. There is one central point that presents the services that are built on top of the platform (Data and AI assets, federated authorization, physical resources, AI tools). These services and these different types of resources can be accessed through the platform itself and through the platform interfaces. At present, two main platform interfaces are available: a REST API documented through an open API protocol and an SDK.

A Python SDK is being developed that communicates directly with the API of the platform. Currently it only exposes a subset of the functionalities provided by the REST API. The visual representation is reported in the following figure (Figure).

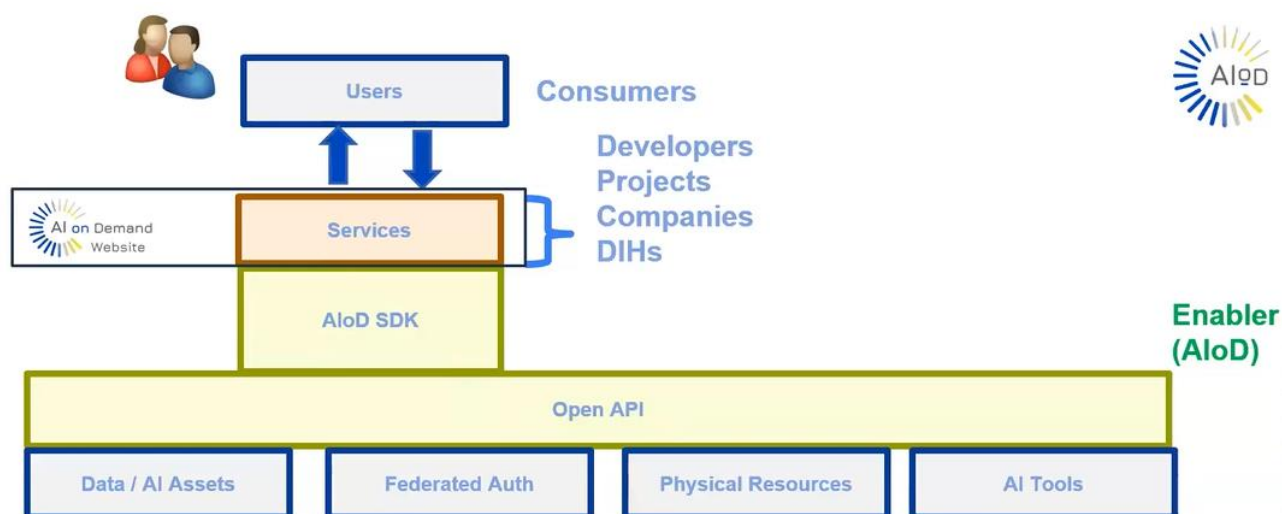


Figure 12 - REST API and Python SDK: integration options for enhanced accessibility

It's important to note that some of these features, such as physical resources and AI tools, are not yet implemented and that the development of the platform is based on successive releases, the sequence of which is illustrated in the next figure, which shows the development stages based on successive releases.

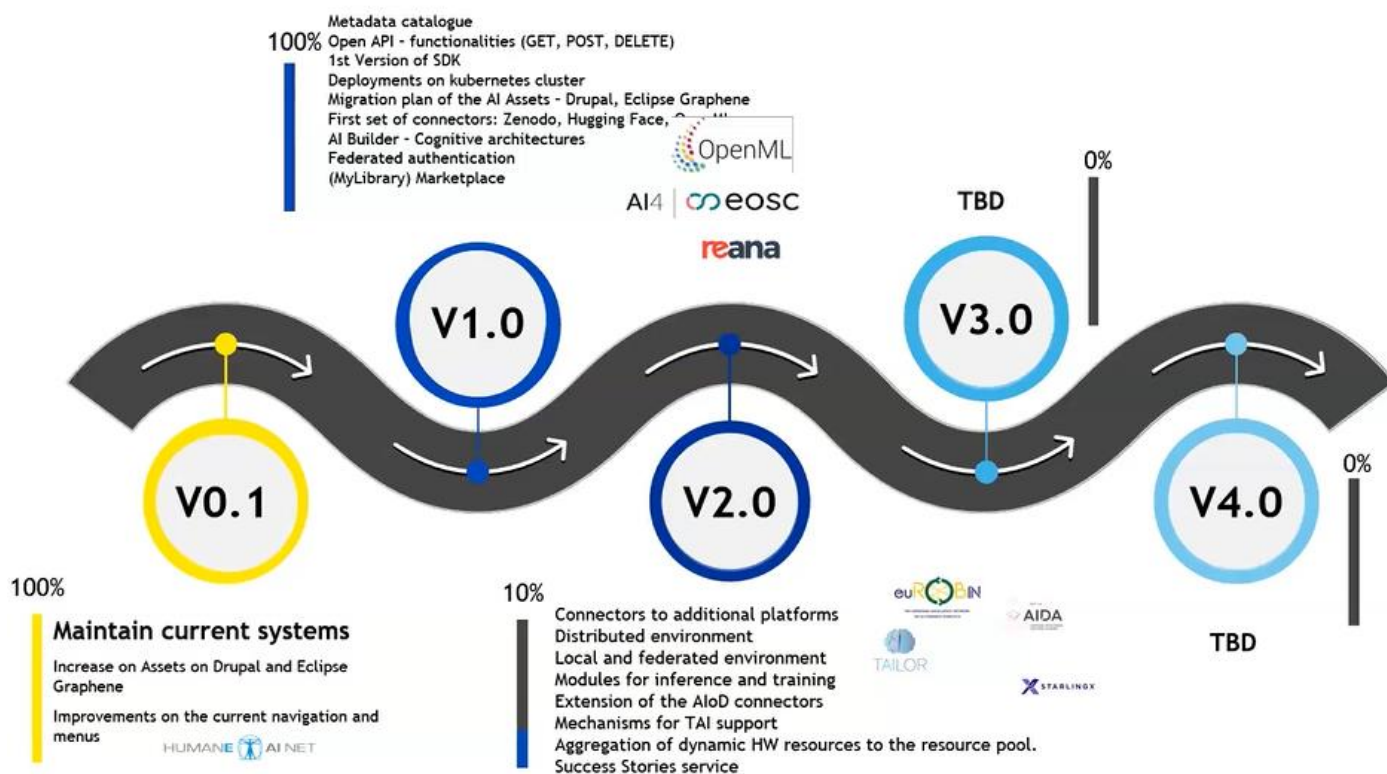


Figure 13 - Development roadmap of AloD platform: implemented features and future releases

At the time of the current state of art analysis, the metadata catalogue provides access to the datasets and the AI resources from three different platforms: Hugging Face, Zenodo and OpenML.

A couple of services are already built. The first one is the **Marketing Portal**. It is the entry point to the platform, that aims to introduce to the platform and to redirect the users to other services built on top of it. This service is focused on marketing and promotion of systems and services, and the UX (User Experience) of this entry point will be improved soon, based on user feedback.

The second one is the **Marketplace**, that aims to allow the user to access different AI Assets on the platform from a UI (User Interface). In the currently available version, it is possible to search for and to access AI models, datasets, experiments and services available in the platform, or rather the metadata that are available in the platform. Furthermore, after searching for assets, it is also possible to access the details of the assets of interest. Users can also create their own collection of assets and services of interest that are added to the personal library (**MyLibrary**). Several datasets can be added to the library, and these will be accessible in other services.

Then there is an authentication and authorization service provided by the platform, which is currently the EGI Check-in. It is based on KEYCLOAK, and it is a federated one which means that one can log in through many different providers.

Finally, there is a service called **RAIL**, which is being built on the platform. This tool is aimed at AI professionals, e.g. researchers, PhD students, but also people from industry. RAIL aims to make it easy to explore and use the AI resources available on the platform. This service enables the combination of AI resources, e.g. datasets and models, and the execution of code on them in a light and flexible way.

The platform can be seen also from the perspective of a service. Interacting with the platform means interacting with the AIoD REST API and its SDK. If one wants to build a service and wants to exploit the platform, the access to various services is provided via On Demand's REST API and its SDK, currently the Python SDK. They provide access to a very rich metadata catalogue containing datasets, AI resources and more, as shown in the following picture. In addition, they provide access to the computational infrastructure through REANA ³(Reproducible research data analysis platform).

³ <https://reanahub.io/>

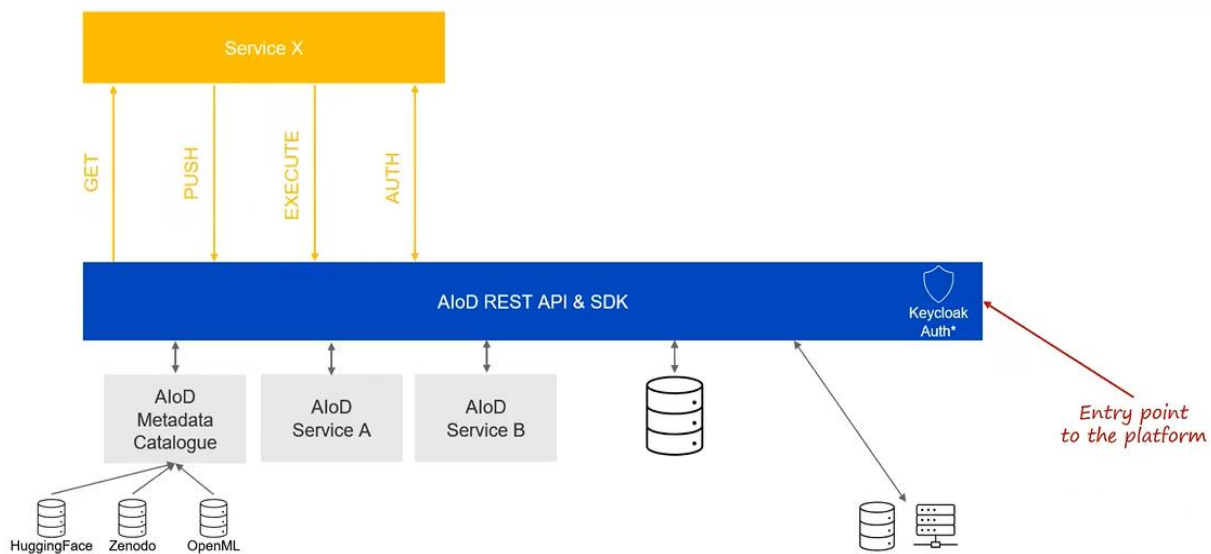


Figure 14 – Interaction with the platform through REST API and SDK

It is important to emphasise that this metadata catalogue does not store the data itself, but the AI-on-Demand platform aims to provide unified access to metadata from not only this platform, but also from other different platforms.

The platform service can basically use its REST API to obtain different resources, it can push resources, it can also execute code in the platform, and it can use the authentication and authorisation mechanism.

The different services can also benefit from each other when using the platform. It is important to note that only the metadata is available in the platform, whereas the data itself is available in the service.

The AI-on-Demand platform is divided into several sections. There is one part of the platform, the **About**⁴ section of the portal that contains information about the AI-on-Demand platform, including: its main features, the main advantages of its use, the purpose of its creation and the main historical stages of its creation. section, in which the main features of the platform and the services it contains are explained.

Then, there is the **oD (on Demand) services** section, which is divided into:

- **Virtual Lab** that contains tools and services to foster the innovation
- **Innovation** that contains tools and services to integrate with AloD

The platform also includes a section **Integrate** that contains developer tools and guidelines to share the AI services. It is divided into two parts:

- **Service Builders**
- **Communication teams**

⁴ <https://aiod.eu/about>

The **Communication teams**⁵ section includes several resources and guidelines to facilitate effective communication and promotion of AI-on-Demand (AloD) initiatives.

Firstly, it includes the Communication Request section, which contains news, announcements, blog articles, publications and press releases. In addition, dissemination and communication guidelines are available, so to further assist the dissemination process. For promotional purposes, is also accessible a range of materials such as brochures, flyers, badges, stickers, pop-up stands and posters, all designed to effectively promote AloD. In addition, Microsoft Office templates, call background images, business cards, folders and e-mail signatures are available to ensure brand consistency across communication channels. A specialised WordPress theme is provided for the development of AloD project websites, which specifically meets their needs. Moreover, the section contains a Brand Book, a comprehensive guide to understanding AloD's brand identity and provides guidance on how to effectively use and maintain brand elements.

Finally, there is the AI **Community**⁶ section that brings together all individuals and organisations interested in contributing to or benefiting from AI capabilities. Researchers, developers, entrepreneurs, citizens and AI experts can engage with AI Ecosystem stakeholders and access multiple resources and solutions. In this part of the platform, it is possible to explore news, events, opportunities about the AloD ecosystem, share the works, the ideas and engaging with the community.

The more technical aspects of the platform are analysed in more detail below.

5.1.1. AloD (on Demand) Services

The section **AloD services**, as also described above, is divided into:

- *Innovation* section which contains tools and services for advanced integration with AloD and companies' examples that demonstrate how AI has helped them expand their business.
- *Virtual Lab* section that contains tools and services that enable researchers to experiment, play and explore, to foster innovation.

The **Innovation**⁷ section is articulated into three subsections:

- *VR Exhibitions*
- *RoboCompass*
- *Success Stories*

The section *Success Stories* of the platform is currently under development and will be available soon, in one of the future planned releases. It will be dedicated to the description of how different companies use AI. Here, each company will have the opportunity to share its own experience.

In the section **VR Exhibitions**⁸ of the portal one can try the Robotics VR exhibition, an immersive experience with interesting technological and robotic solutions from all over Europe, serving the aim of the

⁵ <https://aiod.eu/develop?developTabs=tabList28>

⁶ <https://aiod.eu/services/community>

⁷ <https://aiod.eu/services/innovation>

⁸ <https://aiod.eu/vr-exhibition>

Robotics4EU project: to empower the EU responsible robotics community and ensure a more widespread adoption of (AI-based) robots.

The virtual exhibition was created and developed in cooperation between the EU Robotics4EU project and the Estonian university TalTech. The exhibition can be used on two different platforms: in VR with VR goggles as well as on the webpage, so that people can enjoy the exhibition in 2D on their PC.

To expand the VR exhibition, Robotics4EU is currently looking for companies interested in presenting their robotics solutions in specific sectors. In this way, companies will be able to expand their exhibition and will be able to attract new business partners, reducing the logistical needs or complications that a physical presence at an event would entail for companies.

RoboCompass⁹ (Responsible Robotic Compass) is an innovative tool designed to assess the non-technological aspects of responsible robotics. Particularly, it is an evaluation tool that provides a comprehensive overview of the development of aspects that influence the acceptance of a robot, such as socio-economic issues, human experience, environment, legal issues and data.

To access the service, it is necessary to create an account and log in to this part of the portal. Then, after prioritising according to the importance of each previously listed aspect in the specific case, the user is directed to the dashboard, where the evaluation takes place. An area must be chosen to calculate the overall score. Once the user has clicked on an area, the assessment begins: several questions are asked, aimed to a comprehensive evaluation.

Therefore, *RoboCompass* is an intuitive and easily accessible tool that offers a complete overview of social issues that could affect the use of robots. Moreover, besides performing diagnoses and evaluations, it also provides customised recommendations to improve the user project.

The **Virtual Lab**¹⁰ sub-section of the platform, instead, is itself articulated in several subsections:

- *MyLibrary*
- *AI Builder*
- *Research Bundles (former Platform)*
- *Reproducibility*
- *RAIL*

5.1.2. MyLibrary

All the assets available in the portal are grouped in the **Assets Catalogue**¹¹.

⁹ <https://robocompass.aiod.eu/>

¹⁰ <https://aiod.eu/services/virtual-lab>

¹¹ <https://mylibrary.aiod.eu/marketplace>

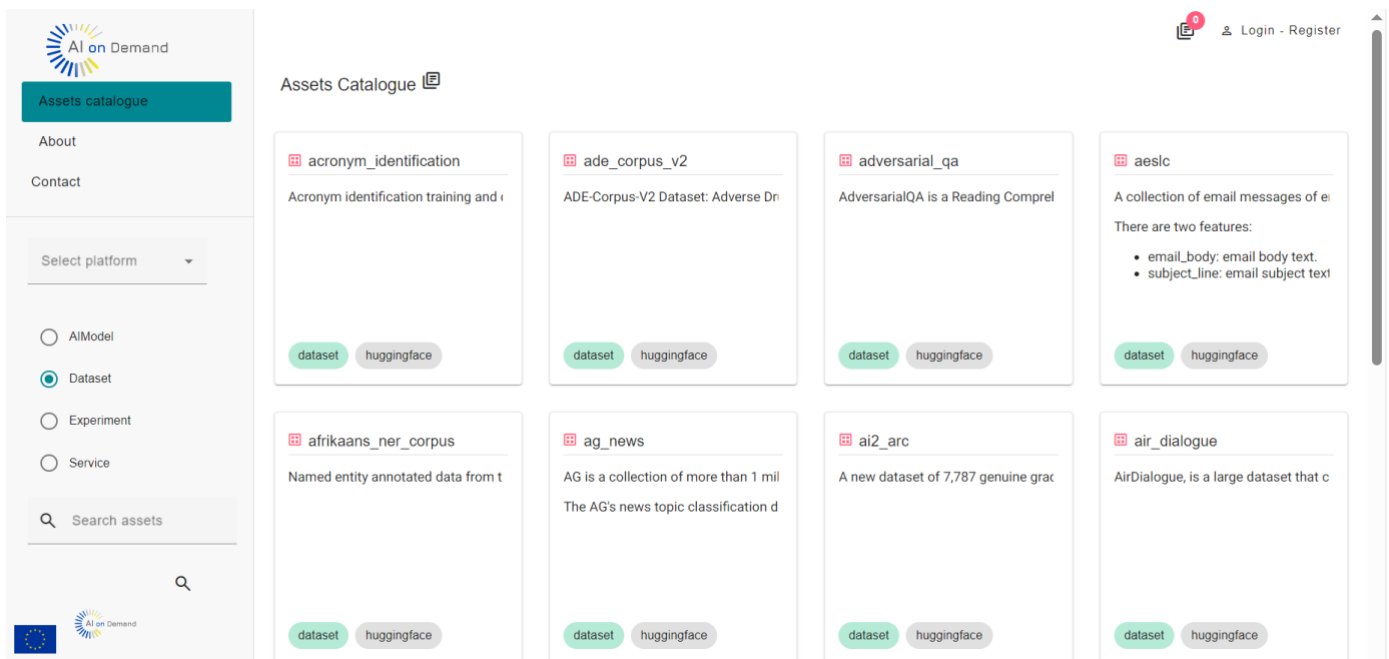


Figure 15 - Discover the available assets in the AIoD Assets Catalogue

MyLibrary is a tool that allows AI researchers to explore some types of AI assets directly in AIoD. It is developed within the AI4Europe project as one of the core services of the new release of the AI-on-Demand platform. More specifically, *MyLibrary* is a web application that enables AI researchers to explore AIoD AI assets, to select the most interesting ones and add them to a personal library that can be shared with other AIoD services. As stated before, *MyLibrary* allows to browse some types of assets in the Assets Catalogue of the AIoD platform, namely: AI models, Datasets, Experiments. The user can filter the assets by the platform or research project they belong to, or search them by keyword, for instance.

Starting from the Assets Catalogue, exploring the contents available on the platform one can click on any asset to learn more about it and, if it is of interest, add it to the Stash on the AI-on-Demand platform. All the assets present in the Stash can be then added to the personal library of AIoD AI assets. For each user the library will be shared with other AIoD services so that each user has the possibility to create the subset of AI Assets of its interest. Thus, it is possible to create a customised library, importing the assets of interest from different services.

It is important to highlight that *MyLibrary* is currently in testing mode.

5.1.3. AI Builder

The **AI Builder**¹² gives access to the on-boarding model¹³, the Marketplace¹⁴, the documentation¹⁵, the list of the personal models (my models¹⁶) and the design studio¹⁷.

¹² <https://aiexp.ai4europe.eu/#/home>

¹³ <https://aiexp.ai4europe.eu/#/modelerResource>

¹⁴ <https://aiexp.ai4europe.eu/#/marketPlace#marketplaceTemplate>

¹⁵ <https://gitlab.eclipse.org/eclipse/graphene/tutorials>

¹⁶ <https://aiexp.ai4europe.eu/#/manageModule>

¹⁷ <https://aiexp.ai4europe.eu/#/designStudio>

The **Marketplace** contains all the available assets present on the platform. They can be filtered following various criteria (by category, by tag or by key words). For each asset, once the user clicks on it, can be seen: a description, the license profile, the signature, the related documents, the model artifacts and the author/publisher details.

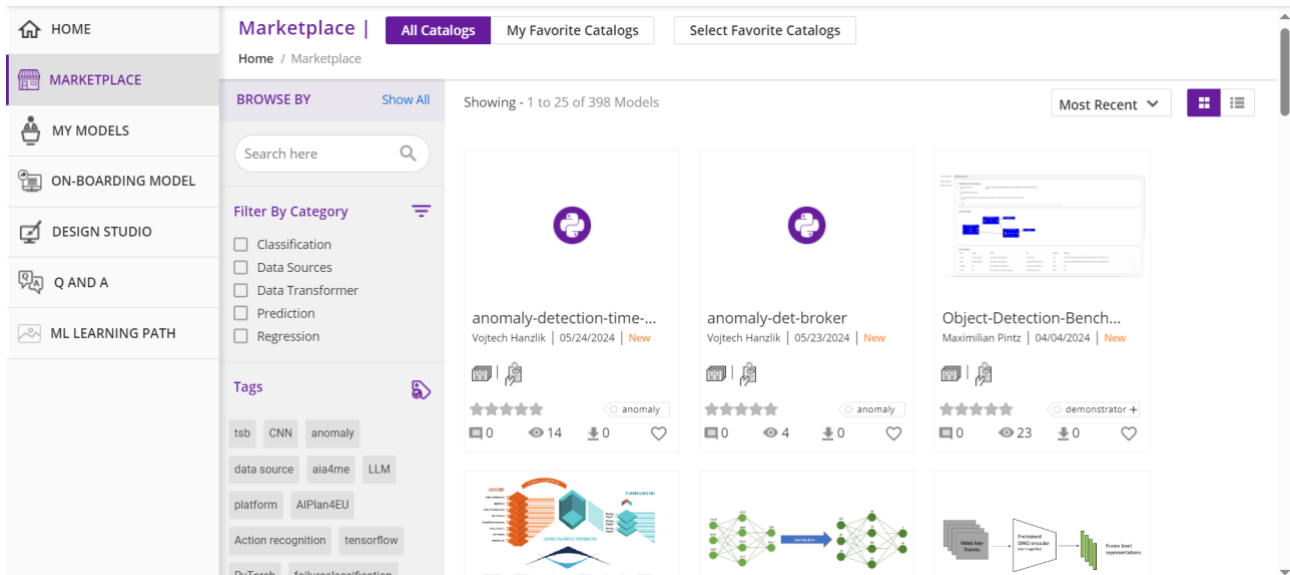


Figure 16 - AIoD Marketplace: explore the available assets

My models section contains the list of the models uploaded by the user, grouped in published, unpublished and deleted models. The models also in this case can be browsed by category or keywords.

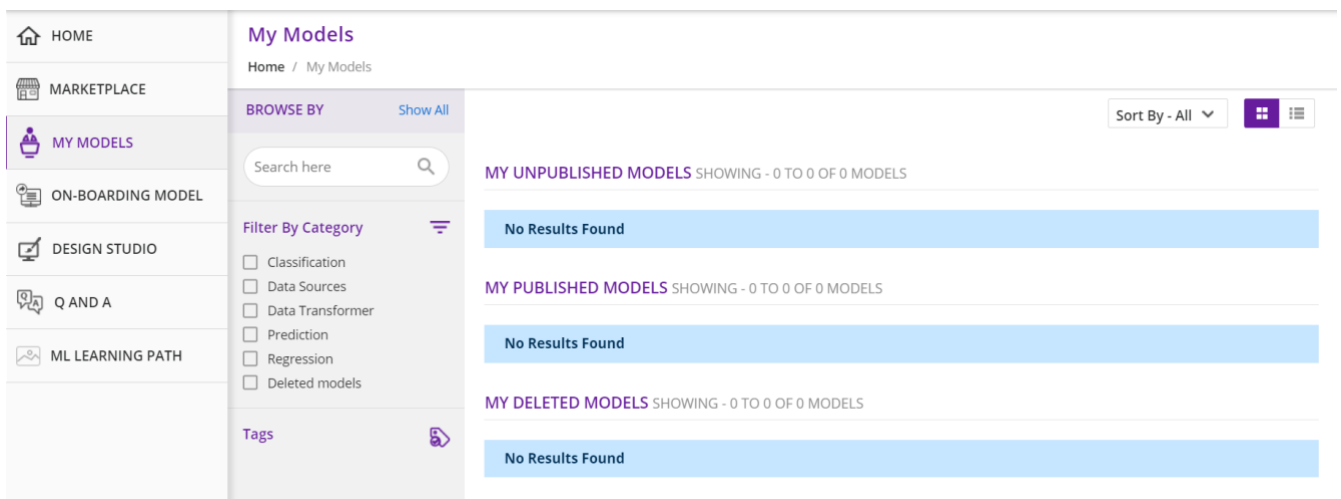


Figure 17 – My Models: a collection of the user's uploaded models

In the **on boarding model** section a template is available to be used in order to upload a model into the platform. Then, there is the **design studio** that allows to create pipelines and composite solutions starting from the available assets. Lastly, it is also available a sub-section containing **Q&A**.

5.1.4. Research Bundles

This section **Research Bundles (former platform)**¹⁸ offers a space in the platform where the results of a research project can be collected and published in a compact manner. A research bundle gathers in one place all the assets (code, data, tutorials, examples, ...) produced by the project and published on the on-demand AI platform. Of course, links to resources published elsewhere, such as GitHub or Zenodo, can be also included. This makes it possible to find and share research content and ideas. Tutorials, documents, code are also available, along with anything else that is deemed useful to the community.

5.1.5. Reproducibility

This feature will allow to create a reproducibility checklist and the workflow to reproduce one's code. Particularly, it will be possible to: ensure that the code is reproducible, evaluate the code against other approaches and upload the code to open-source repositories.

This part of the platform is in development and, therefore, will be available soon, in one of the future programmed releases.

5.1.6. RAIL

RAIL¹⁹ (Research and Innovation AI Lab) is a web application consisting of a backend and a frontend, that allows AI practitioners to explore, search, compare and even create reproducible experiments working with AIoD AI assets. The frontend is coded in a standard technology, Angular, and the backend is also coded in standard technologies, Fast API and Mongo DB. The backend communicates with the REST API of the on-demand platform and accesses resources, models and datasets, especially to read, search and so on. Models, data and execution can be combined. It also uses the authentication and authorisation mechanism of the on-demand platform.

This tool enables to create and execute experiments that can be modified and reused. Then, they can be run directly in the AI-on-Demand platform and on the hardware, infrastructure provided by the platform. Particularly, *RAIL* enables the execution of the assets through REANA (part of AI-on-Demand). If a useful experiment has already been implemented by others it can be reused, while, if the needed experiment is not present in the catalogue, one can implement it and apply it on multiple datasets present in the platform. In addition, it is also possible to download the results of the experiments.

The possibility of running experiments via *RAIL* directly from the Jupiter notebook or from a Python script (RAIL SDK), so that there is no need to use the frontend, is currently under development.

RAIL allows AI practitioners to compose AI assets available in AIoD into executable experiments that are reproducible and reusable. To create an experiment, one must first define the Experiment Template, which is a combination of runtime environment, software dependencies, and code. The Runtime Environment is a specification of the environment in which the experiment will be executed. It includes the specification of the hardware and software environment. The software dependencies are the software packages that are required to run the experiment and, finally, there is the code that is executed in the experiment.

¹⁸ <https://www.ai4europe.eu/research/research-bundles>

¹⁹ <https://rail.aiod.eu/>

An experiment template by itself is not executable, it is necessary to create an experiment based on a specific template to be able to run it. An Experiment is a specific executable instance of an Experiment template. Experiments are composed of AI assets and run in the AloD platform and are reproducible and reusable. The experiment defines certain parameters that are injected into the template via environment variables to make it executable. When an experiment template is created in *RAIL*, a Docker image is built under the hood (after approval), the software dependencies are installed, and the code is packed into the Docker image.

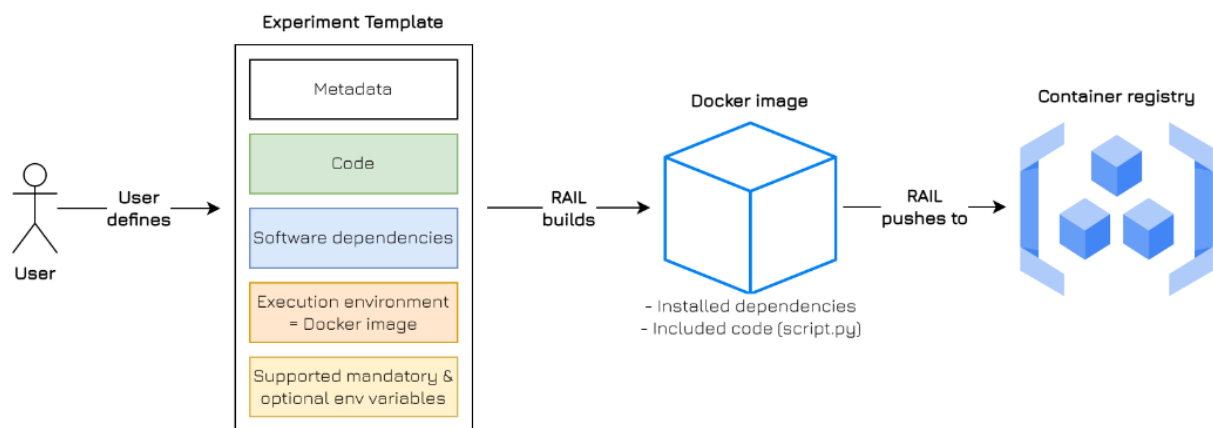


Figure 18 - Process that takes place when an experiment template is created

In the process of creating an experiment, first an Experiment Template, to be use as a basis for the experiment, have to be selected. Then, particular models, datasets and values of environment variables of interest must be selected. An experiment can be created either through RAIL Python SDK or through web UI.

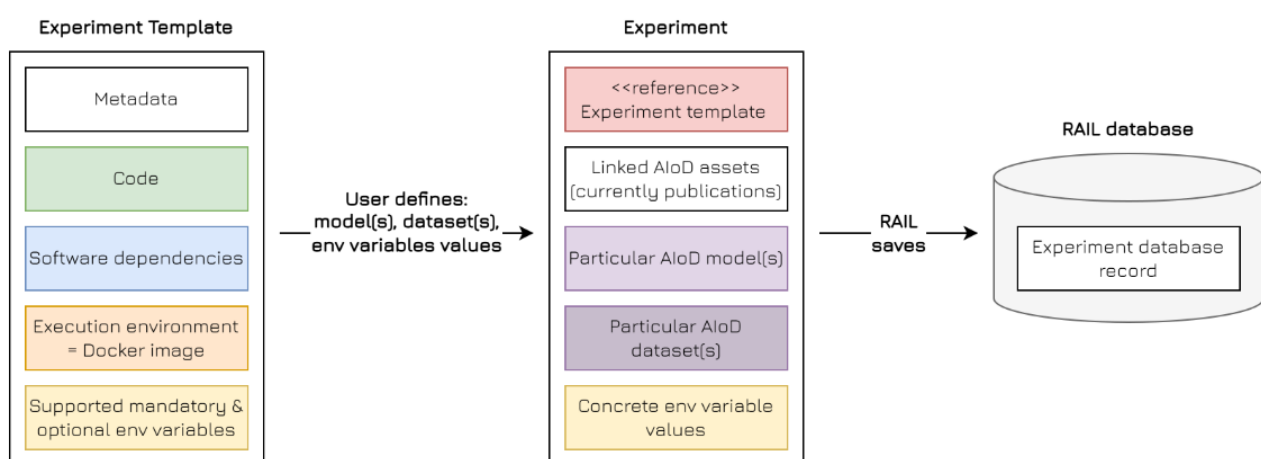


Figure 19 – Experiment creation process

When an experiment is executed after being created, an instance of Experiment Run representing the current state of the running Experiment, its logs, metrics and outputs, is created. Its progress can be

monitored in the Experiment or Experiment Run detail page and through Experiment Run state property. After, the Experiment Run is transformed to a Common Workflow Language (CWL) file that represents the Experiment in the form of a workflow executable by the AIoD platform infrastructure. Then, the CWL file is sent to the AIoD platform's workflow engine, currently implemented as an instance of REANA that executes the Experiment Run. In REANA, a Docker container is started. AI Assets and environment variables are globally available in the container. A file called script.py, which contains the script defined in the Experiment Template, is executed. After the Experiment Run finishes, the Docker logs, metrics and output files, if any, are available on the details of the Experiment Run.

5.1.7. Integrate: Service Builders

The **Service Builders**²⁰ sub-section of the **Integrate**²¹ section, containing tools and guidelines for easy sharing of AI services, is further subdivided into:

- *Learn how to do it*
- *AIoD metadata Catalogue*
- *AIoD open API*
- *Integrate with AIoD*
- *Stylesheet*

The section of the portal **Learn how to do it**²² provides access to a YouTube channel containing some short video tutorials useful to learn how to install, authenticate and access AIoD knowledge and, also, useful to understand how to download and launch AIoD locally.

Accessing the section **Stylesheet**²³ of the portal, useful instruction for saving time on building UIs as well as reusable AIoD guidelines for users UI (User Interface) UX (User Experience) team are available.

In particular, the user is redirected to a Figma page where style sheets for menu, fonts, text styles, colours, cards, favicon, buttons, forms inputs and cookie banner are currently available for two different designs (Light Theme and Dark Mode).

It is important to emphasise that integration is the main theme of the task T5.5 of the project AI REDGIO 5.0, concerning interoperability with the AI-on-demand platform. This task is responsible for ensuring the exposure and interoperability of the AI REDGIO technology/platform portfolio with the AI-on-demand platform. As AI REDGIO 5.0 essentially contributes to the manufacturing sector, several levels of interoperability will be considered, including seamless search and discovery of AI assets from the AI-on-Demand platform; easy publishing and reliable sharing of AI REDGIO 5.0 assets (data, AI models, AI pipeline) in a bidirectional manner (Data layer); and provision and monitoring of experimental facilities to run different AI solutions (Experimentation layer).

²⁰ <https://aiod.eu/develop?developTabs=tabList29>

²¹ <https://aiod.eu/develop>

²² https://www.youtube.com/playlist?list=PLL8OpOdPsmF6hXGJVKT2v_-QUv-FFnwmc

²³ <https://www.figma.com/design/mzR8WbAxxvndvyDCfEXQC0/Stylesheet-AIoD?node-id=1-5894>

The analysis performed on the platform structure is the mandatory starting point for any kind of integration.

5.1.8. AloD metadata Catalogue

The section **AloD metadata Catalogue**²⁴ documents the available APIs useful to interact with the AloD Metadata Catalogue, showing the available endpoints to create, update, delete and list the resources offered on the catalogue.

The APIs on this page are divided in sections, grouped according to the specific asset they allow to manage. For example, there is a section that collects all the methods related to datasets, another section that relates to ML models, and so on. In particular, the first section is about "platforms", entities that provide the assets on the catalogue. These platforms host contents and services that they then decide to showcase on the AloD catalogue; the AloD itself is one of these platforms, and others are the already mentioned OpenML, Hugging Face and Zenodo. The APIs relating to platforms allow to obtain a list of all the registered platforms, and allow an authenticated user to add, remove or modify a platform.

The other sections list all the HTTP methods used to call the API relating to AI assets like datasets, experiments, ML models, but also related to different assets such as educational resources (documentation or video tutorials), computational assets (e.g. storage for data and models or computational clusters), news, contacts and people associated with published assets, and so on. Each asset presents the same set of endpoints: an endpoint allowing to list or count all the assets of one kind (datasets, news, persons...), one to add, modify and delete an asset, one to retrieve assets by their platform of origin, and endpoints to retrieve a single asset given one of its different identifiers.

In addition to the asset-specific endpoints, there is also an endpoint for searching through the catalogue: it allows to search assets by type and to provide a search query to match with the asset's properties.

Moreover, an endpoint exists to list all the available *enums*, that is, all the allowed values for certain tags of the assets: for example, it is possible to retrieve a list of all the application areas an asset covers, or a list of the industrial sectors interested by the asset, or the types of computational assets present.

There are two other endpoints, realistically with the expectation of more to come, that allow a dataset to be uploaded directly to a platform through AloD: this is to allow a registered user to use AloD to both create the item in the catalogue and upload its related content to the platform, rather than having to create the item in the catalogue on the AloD website and then, in a separate operation, visit the platform's website and upload the dataset there. Currently, only the Zenodo and Hugging Face endpoints are available.

5.1.9. AloD open API

The **AloD open API**²⁵ page opens a Swagger UI documentation page that details the same API endpoints described in the previous section 5.3.1.2. In addition to simply listing the endpoints, this page allows the user to test the working API by clicking on each endpoint and making the appropriate HTTP request directly from the documentation page. Particularly, this page also allows the authentication method to be tested, i.e. it

²⁴ <https://api.aiod.eu/redoc>

²⁵ <https://api.aiod.eu/docs>

allows a user to request authentication using one of the OpenID Connect authentication flows supported by AIoD, and to test endpoints that modify the catalogue by adding/modifying or deleting assets.

5.1.10. Integrate with AIoD

The page pointed by the link in this section is a repository hosted on GitHub with the code to deploy an API server instance. The AIoD platform offers the full code as an open-source resource to replicate the same catalogue functionality on a different node. In this repository, apart from the code to run the API server, link to the metadata definitions and API standards are provided.

5.2. Interoperability between the AI REDGIO platform and the AIoD portal

The objective of Task T5.5 of the AI REDGIO 5.0 project is the integration and interoperability of the AI REDGIO portal with the AI-on-demand platform. This task aims to ensure that assets uploaded to the AI REDGIO platform are also available on the AI-on-demand platform. This will facilitate the publication and sharing of AI REDGIO 5.0 assets.

Focusing on the integration task, it is important, once again, to emphasize that the AIoD platform is still in development, thus preventing a full and stable integration with the AI REDGIO portal as of now. The final goal of the integration is providing full collaboration between the AI REDGIO portal and the AIoD, so that users can seamlessly interact with AI-related assets on both platforms.

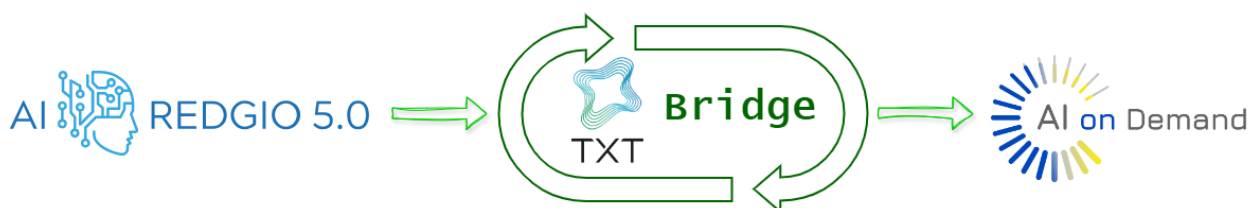


Figure 20 – AI REDGIO portal integration with AIoD Metadata Catalogue

Currently, the only section presenting stable APIs is the AIoD Metadata Catalogue. The project's objective is to implement an interface to make AI assets uploaded on the AI REDGIO portal available also on the AIoD catalogue if desired, with the possibility of expanding the integration in the future. The **Integrate with AIoD**²⁶ section points to a GitHub repository that hosts the code for the REST API server of the AIoD, allowing developers to host a fully functional API server with the full set of endpoints to interact with the catalogue, in order to test integration or build additional services on top of it.

The integration will take the form of a bridge application, which will run in a container separately from the AI REDGIO portal and providing a set of APIs to communicate with the portal. Users creating or managing assets on the AI REDGIO portal will be given the choice, either with a flag option on asset creation or with an appropriate setting on an already created asset, to announce the asset on the AIoD catalogue. At this point the portal will be responsible for contacting the bridge application with the necessary information about the asset, such as its description, its location in the form of a URL, and so on. The bridge application will then use the provided information to create a request according to the AIoD schema and format and create an entry in the AIoD catalogue on behalf of the AI REDGIO portal, thus allowing AIoD users to view the assets present

²⁶ <https://github.com/aiondemand/AIOD-rest-api>

on the portal. Development on the bridge application can be tracked on its public [Github repository](#) within the AI REDGIO organisation.

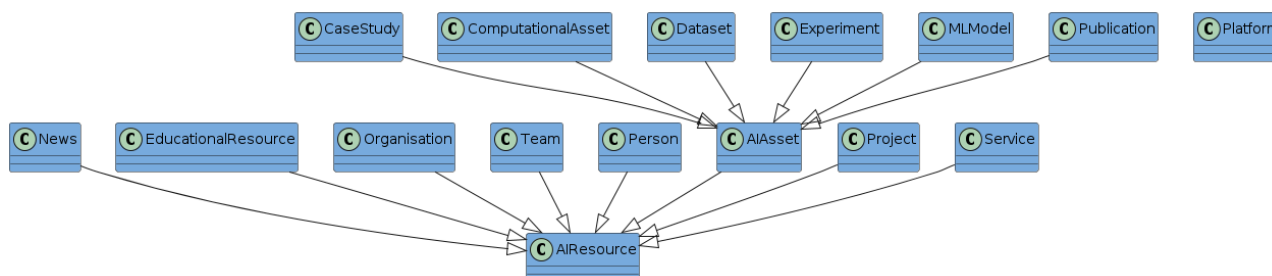


Figure 21 – AIoD available assets Information Model

The bridge will authenticate on the AIoD platform and act on behalf of the AI REDGIO portal after AIoD itself has recognised AI REDGIO as a platform authorised to publish its assets in the Metadata Catalogue. The figure above shows some of the main assets that the AIoD manages on its portal and the relationships between them. When an asset is published, the bridge correctly selects the right type of asset to create in the catalogue.

The following table and figure show the attributes of the base asset, common across all the asset types available in the AIoD catalogue. While the APIs only require the name as mandatory for the creation of an asset, specifying all the attributes allows to better describe what's being published and its characteristic such as its version or if it's connected to other assets.

Table 1 - AI Assets Attributes

AI Asset base attributes	
platform	The external platform from which this resource originates. Leave empty if this item originates from AIoD. If platform is not None, the platform_resource_identifier should be set as well.
platform_resource_identifier	A unique identifier issued by the external platform that's specified in 'platform'. Leave empty if this item is not part of an external platform. For example, for HuggingFace, this should be the /<dataset_name>, and for Openml, the OpenML identifier.
name	The name of this resource.
alternate_name	An alias for the item, commonly used for the resource instead of the name.
distribution	All or part of an AIAsset in downloadable form

keyword	Keywords or tags used to describe this resource, providing additional context.
citation	A bibliographic reference
application_area	The objective of this AI resource
creator	Contact information of persons/organisations that created this resource.
research_area	The research area is similar to the scientific_domain, but more high-level.
same_as	URL of a reference Web page that unambiguously indicates this resource's identity.
is_accessible_for_free	A flag to signal that this asset is accessible at no cost.
version	The version of this asset.
industrial_sector	A business domain where a resource is or can be used.
scientific_domain	The scientific domain is related to the methods with which an objective is reached.
contact	Contact information of persons/organisations that can be contacted about this resource.
date_published	The datetime (utc) on which this resource was first published on an external platform.
has_part	Different items composing the asset.
is_part_of	Identifier of other assets this asset is a part of.
relevant_link	URLs of relevant resources. These resources should not be part of AIoD (use relevant_resource otherwise). This field should only be used if there is no more specific field.
relevant_resource	URLs of relevant resources part of the AIoD related to the asset.

AIAsset
+ platform: string
+ platform_resource_identifier: string
+ name: string
+ alternate_name: string[]
+ description: string
+ distribution: object[]
+ keyword: string[]
+ citation: integer[]
+ application_area: string[]
+ creator: integer[]
+ research_area: string[]
+ same_as: string
+ is_accessible_for_free: boolean
+ version: string
+ industrial_sector: string[]
+ scientific_domain: string[]
+ contact: int[]
+ date_published: string
+ has_part: int[]
+ is_part_of: integer[]
+ relevant_link: string[]
+ relevant_resource: integer[]
+ relevant_to: integer[]

Figure 22 – AI Asset base attributes

6. Conclusion

The main topics of the analysis described in this document are the AI Pipeline Designer platform, the Edge AI reference implementation portfolio, and the AI-on-Demand (AIoD) platform.

The first cited tool facilitates the design, testing, training, execution and deployment of AI models. It also provides the possibility of managing AI pipelines in the cloud and at the edge, integrating federated learning capabilities. It also supports the integration of pre-trained models and pipelines for manufacturing needs.

The second part of the document focuses on the Edge AI reference implementations portfolio, an online wiki in which Edge AI Reference Implementations for the manufacturing domain are published and organised. It includes solutions from AI REDGIO 5.0 partners, open-source resources, and new project-specific knowledge. It serves AI REDGIO partners for their work package activities and is available to general users for implementing Edge AI projects.

Finally, despite the ongoing development of the latest version, an in-depth analysis of the AI-on-Demand (AIoD) platform was reported. In addition to the various features of the platform, the technical tools offered were examined in detail, with particular attention to those that will be of great importance in determining how to proceed with the implementation of the next steps of the project. The knowledge gained from the performed analysis is fundamental for the subsequent development of the project's next phase.

The tools described will be further developed to continuously improve the platform and help creating a more robust and user-friendly environment for developing and deploying AI assets.

The AI Pipeline Designer will be validated through various pilot implementations. Feedback from these pilots will be systematically collected and analysed and will be used to guide the further development and refinement of the AI Pipeline Designer in the second iteration of the project.

The current portfolio of AI assets will be expanded with additional resources to further support the users of the platform. In particular, the next steps for the AI REDGIO 5.0 Portfolio include: (a) increase of the collection of assets by all engaged partners, and in particular the experiments that will have at their disposal more mature results (e.g. models and pipelines), (b) exploration of third-party sources for asset collection based on the mature scenarios of the different experiments, (c) conceptualisation and integration of the User Journeys in the AI REDGIO 5.0 Portfolio wiki, that will facilitate asset exploration and will act as pathways helping users in a step-by-step process to identify the AI REDGIO 5.0 Portfolio assets to achieve their objectives

A detailed strategy to integrate the AI REDGIO portal with the AI-on-Demand platform will also be explored to complete the activities described in Chapter 5. Key findings from the current analysis will serve as the basis for this integration, ensuring interoperability and seamless functionality between the two platforms.

Another core focus will be to explore the strategy for integrating the Pipeline Designer with the AI Builder. New assets, obtained as project results, will be uploaded on the AI-on-Demand platform starting from the AI REDGIO platform, using the bridge described before. This will facilitate the sharing and exploitation of project results and contribute to the growing repository of valuable resources on the platform.