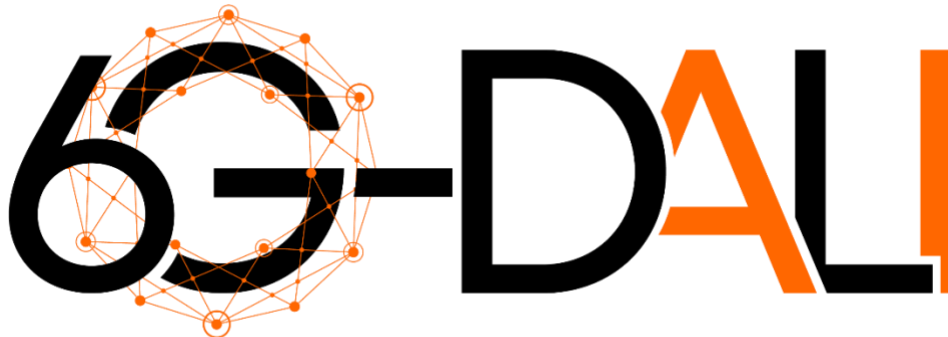


10th September 2025



6G-DALI Business approach and use-cases assessment

Deliverable 2.1, Work Package 2

Project Identifier	HORIZON-JU-SNS-2024-STREAM-B-01-08, Project 101192750		
Project Name	6G Data and ML operations automation via an end-to-end AI framework		
Acronym	6G-DALI		
Start Date	01 January 2025	End Date	31 December 2027
Project URL	www.6gdali.eu		
Deliverable	D2.1 “6G-DALI Business approach and use-cases assessment”		
Work Package	WP2 - 6G-DALI Framework architecture and business model		
Contractual due date	30/06/2025	Actual submission date	10/09/2025
Type	R	Dissemination Level	Public
Lead Beneficiary	OTE		
Responsible Author	Michalis Rantopoulos (OTE)		
Contributors	Panagiotis Marantis (ISI/ATH), Luigi Briguglio (CEL), Francesca Morpurgo (CEL), Lucas Pereira Carwile (CEL), Carmela Occhipinti (CEL), Adlen Ksentini (EUR), Mazene Ameer (EUR), Vasiliki Parousidou (ICOM), Michalis Rantopoulos (OTE), Ioannis Chochliouros (OTE), Giacomo Bernini (NXW), Marco Ruta (NXW), Aitor Alcázar-Fernández (TID), Dimitrios Amaxilatis (SPW), Evdoxia Vasileiou (SPW), Nikolaos Tsironis (SPW)		
Peer reviewer(s)	Vasileios Theodorou (ICOM), Vasiliki Parousidou (ICOM), Aitor Alcázar-Fernández (TID)		
Version	Authors	Comments	
0.1	Michalis Rantopoulos (OTE)	ToC formation	
0.2	Panagiotis Marantis (ISI/ATH), Luigi Briguglio (CEL), Francesca Morpurgo (CEL), Lucas Pereira Carwile (CEL), Carmela Occhipinti (CEL), Adlen Ksentini (EUR), Mazene Ameer (EUR), Vasiliki Parousidou (ICOM), Michalis Rantopoulos (OTE), Giacomo Bernini (NXW), Marco Ruta (NXW), Aitor Alcázar-Fernández (TID), Evdoxia Vasileiou (SPW), Nikolaos Tsironis (SPW)	50% ready	
0.3	Panagiotis Marantis (ISI/ATH), Luigi Briguglio (CEL), Francesca Morpurgo (CEL), Lucas Pereira Carwile (CEL), Carmela Occhipinti (CEL), Adlen Ksentini (EUR), Mazene Ameer (EUR), Vasiliki Parousidou (ICOM), Michalis Rantopoulos (OTE), Giacomo Bernini (NXW), Marco Ruta (NXW), Aitor Alcázar-Fernández (TID), Dimitrios Amaxilatis (SPW)	100% ready	
0.4	Vasileios Theodorou (ICOM), Vasiliki Parousidou (ICOM), Aitor Alcázar-Fernández (TID)	Peer review period	
1.0	Theodora Tsapikouni (ISI/ATH), Christos Verikoukis (ISI/ATH)	QA review	

Disclaimer

The content of this document reflects only the author's view. Neither the European Commission nor the Smart Networks and Services Joint Undertaking (SNS JU) are responsible for any use that may be made of the information it contains.

While the information contained in the documents is believed to be accurate, the authors(s) or any other participant in the 6G-DALI consortium make no warranty of any kind regarding this material including, but not limited to the implied warranties of merchantability and fitness for a particular purpose. Neither the 6G-DALI Consortium nor any of its members, their officers, employees or agents shall be responsible or liable in negligence or otherwise howsoever in respect of any inaccuracy or omission herein.

Without derogating from the generality of the foregoing neither the 6G-DALI Consortium nor any of its members, their officers, employees or agents shall be liable for any direct or indirect or consequential loss or damage caused by or arising from any information advice or inaccuracy or omission herein.

Copyright message

©6G-DALI Consortium. This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both. Reproduction is authorised provided the source is acknowledged.

Table of Contents

Table of Contents.....	4
Glossary of terms and abbreviations used	6
Executive Summary.....	8
1 Introduction	9
1.1 Mapping 6G-DALI Outputs.....	9
1.2 Deliverable Overview and Report Structure.....	10
2 6G-DALI Framework Overview	11
2.1 6G-DALI initial Architecture.....	11
2.2 Description of related frameworks and comparison with 6G-DALI.....	12
2.3 Alignment with other SNS and HE projects.....	15
2.3.1 SUNRISE-6G.....	15
2.3.2 6G-BRICKS	15
2.3.3 6G-SANDBOX	16
2.3.4 ADROIT6G	16
2.3.5 6G4SOCIETY.....	17
3 Business Analysis.....	18
3.1 Market Analysis and Opportunities	18
3.1.1 Market Analysis for Data Spaces.....	18
3.1.2 Market Analysis for 6G Networks and Telecom Data	18
3.1.3 Market Analysis and Potential for AI-as-a-Service (AlaaS) in 6G-DALI.....	19
3.1.4 Data Scarcity in the Telecom domain.....	21
3.1.5 Strategic Opportunity in Addressing the 6G Data Gap	22
3.1.6 6G-DALI's Strategic Response.....	23
3.2 6G-DALI's Stakeholders and Impact.....	24
3.2.1 Identification of the 6G-DALI stakeholders and their roles	24
3.2.2 6G-DALI Impact.....	25
3.2.3 6G-DALI Value Proposition to Stakeholders.....	26
3.3 Business Models	27
3.3.1 Gaia-X Business Model.....	27
3.3.2 6G-DALI Business Model	29
3.3.3 Alignment with Gaia-X principles.....	31
4 Description of the PoCs.....	33
4.1 PoC#1 - Data management and experiment on demand	33
4.1.1 Description	33
4.1.2 PoC participants	35
4.1.3 KPIs Definitions	36
4.1.4 KVis definitions	36
4.2 PoC#2 - AlaaS for CDN apps via cross-testbed decentralized MLOps	39
4.2.1 Description	39
4.2.2 PoC participants	40
4.2.3 KPIs Definitions	41
4.2.4 KVis Definitions.....	41
4.3 PoC#3 - DTT and RLOps for large and medium scale experiments.....	44
4.3.1 Description	44
4.3.2 PoC participants	46
4.3.3 KPIs Definitions	46
4.3.4 KVis definitions	47
4.4 Sustainability Analysis.....	48
4.5 Requirements Elicitation Process.....	49
4.5.1 Personae profiles and user stories for each PoC.....	50

5	Conclusions	56
6	References	57

List of Figures

Figure 1: 6G-DALI AI framework - high-level architecture.....	11
Figure 2: Relationship between 6G-DALI's stakeholders	25
Figure 3: DataOps for existing and on demand data sets.	34

List of Tables

Table 1: Adherence to 6G-DALI GA Deliverable & Tasks Descriptions.....	9
Table 2: Comparison of 6G-DALI with other frameworks.....	14
Table 3: Business Model Types enabled by Gaia-X.....	28
Table 4: Business Model Canvas 6G-DALI.....	32
Table 5: 6G-DALI framework quantitative evaluation of Sustainability	48
Table 6: General use cases requirements	49
Table 7: Personae profiles for PoC#1	50
Table 8: User Story 1 - Data set Discovery Using Natural Language in PoC#1	51
Table 9: User Story 2 - Triggering ETL Pipeline for Existing Data sets in PoC#1	51
Table 10: User Story 3 - Generating New Data set via Experiment Orchestration in PoC#1.....	51
Table 11: User Story 4 - ELT Pipeline Execution for Generated Data in PoC#1	51
Table 12: Personae profiles for PoC#2	52
Table 13: User Story 5 - Model Compression in PoC#2.....	53
Table 14: User Story 6 - Model Benchmarking in PoC#2.....	53
Table 15: User Story 7- Model Drift Detection / Continuous Monitoring in PoC#2	53
Table 16: Personae profiles for PoC#3	54
Table 17: User Story 8 - Automated Reinforcement Learning Agent Validation in PoC#3	54
Table 18: User Story 9 - Data set generation for Mobility Optimisation in PoC#3.....	54
Table 19: User Story 10 - Real-Time Monitoring and Adaptive Experimentation in PoC#3.....	55
Table 20: User Story 11 - High-Fidelity O-RAN Scenario Testing in PoC#3	55

Glossary of terms and abbreviations used

Abbreviation / Term	Description
AEMD	AI Experts and ML Developers
AI	Artificial Intelligence
AlaaS	AI-as-a-Service
APIs	Application Programming Interfaces
BMC	Business Model Canvas
CEDS	Common European Data Spaces
CDN	Content Delivery Network
CISP	Cloud Infrastructure Service Provider
CN	Core Network
COOPERANTS	Collaborative Processes and Services for Aeronautics and Space
DaaS	Data-as-a-Service
DataOps	Data Operations
DP	Data Provider
DPaaS	Data Platform as a Service Provider
DT	Digital Twin
DTT	Digital Twin Testbed
E2E	End To End
ELT	Extract Load Transform
ETL	Extract Transform Load
EOSC	European Open Science Cloud
ETHAI	Ethics of AI
ETSI	European Telecommunications Standards Institute
FAIR	Findable-Accessible-Interoperable-Reusable
FLOps	Federated Learning Operations
FP32	32-bit floating point
FSP	Framework Service Provider
GDPR	General Data Protection Regulation
GoKVI	Groups of Key Value Indicators
GUI	Graphical User Interface
HPO	Hyperparameter Optimization
IDC	International Data Corporation
IDSA	International Data Spaces Association
IoT	Internet of Things
ITU	International Telecommunication Union
JCAS	Joint Communication and Sensing
KPI	Key Performance Indicator
KSI	Key Sustainability Indicators
KV	Key Values
KVI	Key Value Indicator
LCI	Legal Compliance Index
LLM	Large Language Model
ML	Machine Learning
MLOps	Machine Learning Operations
MVNOs	Mobile Virtual Network Operators
NaaS	Network-as-a-service
NSaaS	Network Slice-as-a-service

NCP	Network Cloud Provider
ORAN	Open Radio Access Network
PHY	Physical Layer
PoC	Proof-of-Concept
QoE	Quality of Experience
RIS	Reconfigurable Intelligent Surfaces
RLOps	Reinforcement Learning Operations
RRI	Responsible Research and Innovation
SAT	Social Acceptance of Technology
SDGs	Sustainable Development Goals
SIM2REAL	Simulation-to-Reality
SLA	Service Level Agreement
SME	Small and Medium-sized Enterprise
SNS JU	Smart Networks and Services Joint Undertaking
SNVC	Societal Needs and Value Creation
SSH	Social Sciences and Humanities
SW	Software
TRL	Technology Readiness Level
UE	User Equipment
VM	Virtual Machines
vRAN	Virtualised Radio Access Network
VSP	Vertical Service Provider

Executive Summary

This deliverable, D2.1 “6G-DALI Business approach and use-cases assessment”, forms a critical part of the 6G-DALI project (Project Identifier: HORIZON-JU-SNS-2024-STREAM-B-01-08, Project 101192750), a pioneering initiative within the Horizon Europe programme that aims to promote the 6G Data usage and ML operations automation via an end-to-end AI framework. This report meticulously details the business approach and comprehensively assesses the project’s Proofs of Concept (PoCs), including the definition of Key Performance Indicators (KPIs) and Key Value Indicators (KVIIs).

The 6G-DALI project addresses the crucial need for an end-to-end Artificial Intelligence (AI) framework to support the evolution of 6G networks. At its core, it introduces a novel architecture built upon two synergistic pillars: AI experimentation as a service via Machine Learning Operations (MLOps) and data collection and storage via Data Operations (DataOps). This framework is designed to automate MLOps and DataOps workflows, fostering a robust 6G data economy by enabling secure, trusted, and sovereign data exchange among diverse stakeholders, aligned with International Data Spaces (IDS) and Gaia-X principles. The framework’s unique integration of AI-driven experimentation strategically places 6G-DALI in a complementary and synergistic position regarding prominent data governance initiatives and Gaia-X compliant projects, creating an active environment for AI deployment and data generation on demand. Further details on this foundational framework and its comparison with other initiatives can be found in Section 2.

In Section 3 a thorough business analysis identifies significant market opportunities, particularly in addressing the prevalent data scarcity within the telecommunications domain. The project’s strategic response is underpinned by a comprehensive Business Model Canvas (BMC), which elaborates the monetisation opportunities, key stakeholders, and the distinctive value proposition 6G-DALI offers, thereby contributing to the development of a Common European Data Space (CEDS) aligned with European values and strategy.

The technical feasibility and key components of the 6G-DALI solution are validated through three distinct Proofs of Concept (PoCs). Section 4, “Description of the PoCs”, provides an in-depth outline of each PoC, including their descriptions, participants’ needs and benefits, associated KPIs and KVIIs. These PoCs demonstrate advanced capabilities such as data management and experiment on demand, AI-as-a-Service (AIaaS) for Content Delivery Network (CDN) applications via cross-testbed decentralised MLOps, and Digital Twin Testbed (DTT) and Reinforcement Learning Operations (RLOps) for large and medium-scale experiments. In addition to the above, a sustainability analysis of the e2e AI framework under societal, economic and environmental aspects, is performed based on the 3 PoCs outcomes.

The multifaceted approach of 6G-DALI is poised to significantly impact the European technology landscape, strengthening digital sovereignty and fostering a competitive European data economy by accelerating AI model development, reducing operational costs, and enhancing service quality. Through its strategic alignment with European 6G initiatives and its commitment to ethical and legal frameworks, 6G-DALI serves as a foundational enabler for the entire 6G AI ecosystem, empowering stakeholders to innovate with confidence and in adherence to European values.

1 Introduction

This section provides a brief outline of the objectives of the specific 6G-DALI Deliverable, how those are aligned and relevant with the overall project, and what was the approach followed in order to achieve them.

1.1 Mapping 6G-DALI Outputs

The purpose of this section is to map 6G-DALI Grant Agreement commitments, both within the formal Deliverable and Task description, against the project's respective outputs and work performed (see Table 1).

Table 1: Adherence to 6G-DALI GA Deliverable & Tasks Descriptions

Deliverable description			
D2.1 '6G-DALI Business approach and use-cases assessment' - Report on the business approach of 6G-DALI, refinement, and assessment of the PoCs, including KPIs and KVIs			
6G-DALI Task No. and Title	6G-DALI GA Task Description	Respective Document Chapter(s)	Justification
Task 2.1 Business analysis and use-case requirements	6G-DALI's exposure via Gaia-X will help to organize the community interested in contributing data sets to the 6G data space, such as future and ongoing SNS projects as well as CL4 Horizon program projects.	Section 2 – 6G-DALI Framework Overview	Section 2 provides a comprehensive overview of the 6G-DALI project, along with comparison and alignment with other large-scale initiatives and research projects. Additionally, we compare 6G-DALI with other prominent data governance initiatives and Gaia-X compliant projects, highlighting its unique integration of MLOps and AI-driven experimentation. We outline 6G-DALI's strategic alignment with other significant European 6G initiatives and Horizon Europe projects.
	The objective of this task is to identify the opportunities to monetise the generated data sets through the Gaia-X eco-systems to enable wide exposure of the collected data sets on 6G to the community, including all stakeholders: ML developers, Verticals, Network Operators, and manufacturers (particularly the new ones working	Section 3 – Business Analysis	This section describes thoroughly the monetisation opportunities of 6G-DALI. It highlights in great detail the reasons for the so-called "Data Scarcity" telecom data sets and focuses on the strategic opportunity that this creates for the project. Additionally, it defines all the relevant stakeholders, and the value proposition 6G-DALI creates for them. A comprehensive Business Model Canvas (BMC) has been developed to serve as

	on Open RAN).		a complete representation of the 6G-DALI business model
	The task will reassess the pre-defined set of use cases and perform adjustments as seen fit to adjust with technological advances and market evolution and ensure feasibility and relevance. Finally, this task will revisit the determined KPIs.	Section 4 – Description of the PoCs	This Section outlines the three Proofs of Concept (PoCs) that will be developed during the project to validate the key components of the overall solution and demonstrate its technical feasibility. Apart from the description of the three PoCs, their participants' needs, their benefits as well as the corresponding KPIs and KVIs are provided. In addition to the above, a sustainability analysis of the e2e AI framework is performed based on the 3 PoCs outcomes.

1.2 Deliverable Overview and Report Structure

This document consists of the following chapters which may be summarized as below:

- **Chapter 1:** Introductory part to the document with clear identification of the Chapters' contents.
- **Chapter 2:** This part of the document gives initially, a high-level overview of the 6G-DALI architecture, so that the reader gets familiar with the MLOps and DataOps pillars. Furthermore, 6G-DALI is compared with other prominent data governance initiatives and Gaia-X compliant projects, showcasing the innovative advantage of the project.
- **Chapter 3:** Provides a complete Business Analysis. The monetisation opportunities of 6G-DALI are clearly identified in this section and the strategic opportunities due to Data Scarcity. A Business Model Canvas is the strategic tool chosen to provide a consolidated overview of the developed business model.
- **Chapter 4:** Provides a complete description of the three PoCs that will be developed during the project. The associated KPIs and KVIs are also defined, as well as a sustainability analysis of the e2e AI framework. In addition to the above, the user requirements are elicited by means of relevant user stories.

2 6G-DALI Framework Overview

This section provides a comprehensive overview of the 6G-DALI project, along with comparison and alignment with other large-scale initiatives and research projects. Specifically, Section 2.1 outlines 6G-DALI's high-level architecture and its key objectives focused on an end-to-end DataOps and MLOps framework for 6G. In Section 2.2 we compare 6G-DALI with other prominent data governance initiatives and Gaia-X compliant projects, highlighting its unique integration of MLOps and AI-driven experimentation. Finally, Section 2.3 outlines 6G-DALI's strategic alignment with other significant European 6G initiatives and Horizon Europe projects.

2.1 6G-DALI initial Architecture

The high-level architecture of 6G-DALI, depicted in Figure 1, outlines the main components of the end-to-end AI framework. It is designed around two main pillars: **(1) AI experimentation as a service via MLOps** and **(2) Data and analytics collection and storage via DataOps**, both supporting the automation of MLOps and DataOps workflows, respectively. The framework serves both ML developers, focused on optimising 6G infrastructure and enabling closed-loop, zero touch management, and vertical users developing ML models to enhance their 6G services. The framework also interfaces with a range of 6G-testbeds, encompassing RAN, core networks (CN), and computing infrastructure (from edge to cloud), with some also including extreme-edge devices. Moreover, there is at least one Digital Twin (DT) for large-scale testing. Central to this architecture is the creation of a 6G Data Space that links data producers and consumers, enabling data set generation, referencing, and sharing. Relied on the IDSA¹ and Gaia-X principles, this Dataspace promotes secure, trusted, and sovereign data exchange among stakeholders (ML developers, verticals, operators, and vendors) paving the way for a robust 6G data economy.

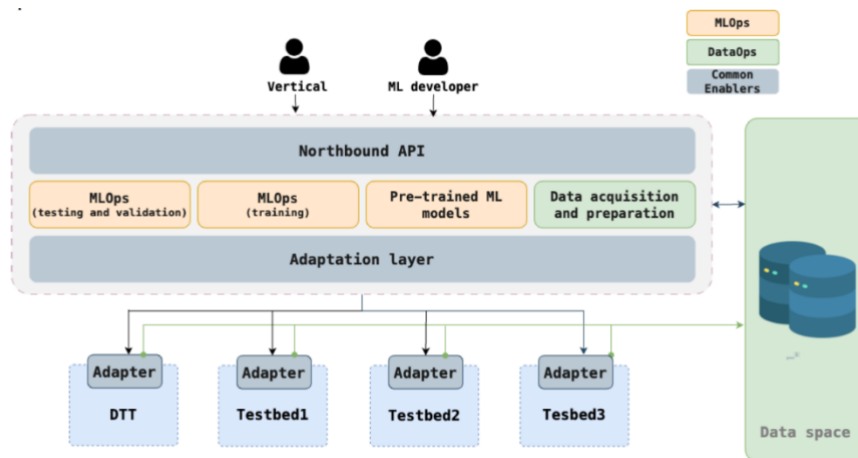


Figure 1: 6G-DALI AI framework - high-level architecture

6G-DALI defines seven measurable objectives which, when combined, contribute to the materialization of the two Pillars of the project:

- **Objective 1: Devise a user-friendly e2e AI framework for DataOps and MLOps.** The first objective focuses on developing an intuitive end-to-end AI framework for 6G, built around integrated MLOps and DataOps pillars. It aims to provide high-level graphical and API-based interfaces for seamless interaction, enabling AI experimentation as a service, intent-driven data handling, and interaction with Large Language Models (LLMs). Users will be able to discover data sets, request data sets via intents, improve data quality, test ML models, and validate them on testbeds, including a Digital Twin.
- **Objective 2: Deliver Gaia-X and Extract Load Transform (ELT) approaches for DataOps in 6G environments.** 6G-DALI introduces mechanisms for handling both “cold data” (pre-existing data

¹ International Data Spaces Association: Available at: <https://internationaldata spaces.org/>

sets discoverable through secure Gaia-X catalogues) and “hot data” generated in real time via 6G testbeds and the digital twin testbed. Enhanced ELT pipelines, featuring advanced data cleaning and augmentation methods, will ensure high-quality data sets for ML model training. Moreover, the 6G-DALI DataOps framework will be designed for reuse across other SNS projects, promoting interoperability and contributing to a reference 6G data space.

- **Objective 3: Streamline trustworthy AI/ML operations via MLOps and AutoML.** 6G-DALI will support the full AI model lifecycle, from training and validation to performance benchmarking and deployment, including also techniques such as Reinforcement Learning Operations (RLOps) and Federated Learning Operations (FLOps), integrated with 6G-testbeds and Digital Twins. A meta-orchestration layer will also unify diverse MLOps/RLOps stacks, offering pre-trained models, validations tools, and mechanisms to assess model accuracy, performance, and trustworthiness, thereby delivering AI-as-a-Service (AIaaS) for 6G Experimentation.
- **Objective 4: Deliver plug-able adapters to easily integrate 6G testbeds.** 6G-DALI will develop plug-able adapters and an adaptation layer that standardizes MLOps and DataOps Application Programming Interfaces (APIs) to ensure seamless integration of both current and future 6G testbeds into the 6G-DALI AI framework. These components will translate between the AI framework’s northbound interfaces and the specific APIs of the testbeds and Digital Twins, enabling easy onboarding of new testbeds (e.g., those from future SNS projects).
- **Objective 5: Build and integrate a Digital Twin Testbed (DTT) to generate representative 6G data sets.** This objective focuses on developing a high-fidelity Digital Twin Testbed to complement existing 6G testbeds by enabling scalable, automated experimentation and data set generation. The DTT will simulate large-scale scenarios, such as massive User Equipment (UE) connectivity or multi-cell mobility testing, that are challenging to replicate in physical testbeds. Built on OpenAirInterface (OAI) and VIAVI’s tools, and following an Open RAN architecture (CU, DU, RU split), the DTT will support automated test execution and data collection to produce realistic 6G data sets.
- **Objective 6: Ethical data sets and legislative compliance.** This objective ensures that ethical considerations and legal compliance are integrated into the development and use of data sets and validation methods within the 6G-DALI framework. The project will align with current regulations such as GDPR, the Data Governance Act, the Data Act, and the AI Act to foster trust and societal acceptance.
- **Objective 7: Dissemination, Communication, Exploitation and Standardization.** 6G-DALI’s impact will be maximized through dissemination, communication, exploitation and standardization activities. This project will generate a portfolio of technological innovations to be shared via scientific publications, workshops, and contributions to key standard bodies like 3GPP, IRTF, ETSI, and ITU, including work on AI Ethics. It will also actively engage with open-source communities and collaborative initiatives such as the Linux Foundation, OpenAirInterface², ZenML,³ SNS JU, and Gaia-X.

2.2 Description of related frameworks and comparison with 6G-DALI

Fuelled by this digital transformation era and the need for effective data governance, various global or European initiatives, projects and frameworks have emerged designed to foster sovereign, interoperable, and trustworthy data exchange across sectors. These frameworks, similarly to 6G-DALI, aim to offer a standardized and effective approach to sovereign and secure data exchange and monetization in domains such as healthcare, energy, manufacturing, and telecommunications, with Gaia-X being a notable European initiative in this direction. In this section we present an overview of prominent frameworks and Gaia-X-compliant projects. We also provide a detailed comparison with the 6G-DALI project, which introduces a novel approach by tightly integrating MLOps and AI-driven experimentation in the 6G domain.

² OpenAirInterface: Available at: <https://openairinterface.org/>

³ ZenML: Available at: <https://www.zenml.io/>

First, we provide an overview of three large-scale initiatives focused around defining principles and standards for data sharing and governance in the fields of both industry and research:

- **International Data Spaces Association (IDSA):** The IDSA provides an architecture for secure, sovereign, and standardized data exchange across industries. Its central concept is enabling organizations to maintain full control over their data. IDSA recently focused on also unifying their concept architecture with the Telecommunications domain [1], by highlighting the important role of operators in providing the foundational infrastructure required for secure data exchange. While IDSA is highly influential, it primarily focuses on defining frameworks and standards rather than implementing specific technical solutions or MLOps capabilities.
- **European Open Science Cloud (EOSC):** The European Open Science Cloud⁴ (EOSC) is a European Commission initiative designed to provide an open environment for research data and services based on the FAIR (Findable-Accessible-Interoperable-Reusable) [2] principles for scientific data management. The EOSC infrastructure implementation is directly supported by several Horizon Europe projects. EOSC is also involved in fostering machine learning and AI-driven research through open data availability and compute environments, making it highly relevant to MLOps in science and research domains.
- **Gaia-X:** Gaia-X⁵ is a European initiative aiming to develop a federated, open data infrastructure based on transparency, data sovereignty, and privacy. It provides specifications and requirements to ensure compliance with European legislation and ethical standards. Gaia-X does not offer ready-made technical solutions, but it serves as a meta-framework, guiding the development of domain-specific dataspace projects. 6G-DALI aims to be fully compliant with the Gaia-X specifications, adapting them into the 6G domain. In Section 3 we also provide a comprehensive analysis of the Gaia-X principles and business model.

We proceed with outlining more specific Gaia-X compliant projects or initiatives applied to various vertical and technological domains, which we also compare against 6G-DALI:

1. **Gaia-X Collaborative Processes and Services for Aeronautics and Space (COOPERANTS):** The Cooperants⁶ project is a German initiative that applies the Gaia-X principles into usable technologies by building data infrastructures mainly focused on the aerospace sector. It partially supports AI integration and MLOps workflows and aims to provide functional digital twin environments to virtually represent robotic systems and spacecrafts.
2. **Gaia-X DataSpace4Health:** Dataspace4Health⁷ is a healthcare-specific implementation of Gaia-X principles, focusing on enabling privacy-preserving and interoperable data sharing across hospitals, labs, and researchers. By creating a secure health dataspace with Gaia-X compliant services, the project aims to securely offer health data to researchers while preserving data ownership and privacy. The project's infrastructure aims to support federated access control and consent management, promoting cross-border collaboration and evidence-based medical innovation. While the project does explore AI solutions for medical treatments, it is more focused on infrastructure and compliance than on directly integrating advanced MLOps workflows.
3. **Omega-X:** Omega-X⁸ is a Horizon Europe project that applies IDSA and Gaia-X standards to the energy sector. It aims to implement an energy data space containing data available but currently underused in energy systems, while also including federated infrastructure, data marketplace and service marketplace. It supports data sharing between different stakeholders while guaranteeing scalability and interoperability with other data space initiatives. The goal is to support data-driven energy use-cases and related services such as grid optimization, renewable integration, and dynamic

⁴ <https://eosc.eu/>

⁵ <https://Gaia-X.eu/>

⁶ <https://cooperants.de/>

⁷ <https://www.dataspace4health.lu/>

⁸ <https://omega-x.eu/>

- pricing. Nevertheless, while it touches on advanced use cases, it is not centered around AI and MLOps.
4. **Catena-X**: Catena-X⁹ is an open and collaborative data ecosystem designed for the automotive sector, with industry members across the globe. It implements both Gaia-X and IDSA principles to facilitate sovereign, standardized and trusted data exchange between manufacturers, suppliers and Small and Medium-sized enterprises (SMEs). Key applications include resilient supply chains through an exchange of demand forecasts and production capacity data, data-driven predictive maintenance, emissions tracking, and quality control. Catena-X stands out for its wide-scale industrial deployment, but is primarily focused on data governance and exchange rather than machine learning or AI workflows.
 5. **Pontus-X**: Pontus-X¹⁰ is a decentralized data and service exchange ecosystem that complies with Gaia-X governance principles. It supports federated learning, federated analytics, and secure data sharing among organizations. Pontus-X enables participants to exchange software, data, and AI artifacts with full sovereignty and monetization options. Its implementation is advanced, including a portal and a Graphical User Interface (GUI) that allows all participants to publish or consume a multitude of available industrial data sets, AI algorithms and Software-as-a-Service solutions in the active available catalogue.
 6. **Flex4Res**: Flex4Res¹¹ is a Horizon Europe-funded project aimed at strengthening the resilience of European manufacturing and supply chains. It combines the Gaia-X and IDSA frameworks with Asset Administration Shell technology to build an open platform that facilitates data sharing across the manufacturing chain. It also incorporates two main toolboxes, one for resilience assessment and another dedicated to optimizing reconfiguration strategies through data monitoring and prediction. Finally, a digital twin of the manufacturing network is also central to the project for testing manufacturing processes before they are implemented; nevertheless, AI integration is limited to only one of the use-cases.

Similarly to the described projects, 6G-DALI applies Gaia-X principles in data governance and DataOps to a specific application context, the **telecommunications domain**. Table 2 provides a direct comparison of 6G-DALI with the other frameworks.

6G-DALI differentiates itself by enhancing the Gaia-X high-level framework with direct MLOps integration and data generation on demand, enabled by a digital twin. These capabilities are crucial for AI model development and testing in the 6G domain, where real-time data may be scarce, dynamic, or privacy sensitive. While many other Gaia-X-compliant projects focus on enabling secure data sharing or data catalogues, 6G-DALI further advances by creating an active experimentation and AI deployment environment.

Table 2: Comparison of 6G-DALI with other frameworks

Initiative-Project-Framework	Domain	Gaia-X compliant	Digital Twin integration	MLOps/AI integration	Data-on-Demand capability
Large Scale reference Initiatives and Frameworks					
IDSA	Cross-Domain	Compatible	Indirect	Supported	No
EOSC	Research	Partially Compatible	No	Moderate	No
Gaia-X	Cross-Domain	Reference Framework	Indirect	Supported	No
Gaia-X compliant projects applied to specific technological domains					
Cooperants	Aerospace Engineering	Yes	Yes	Moderate	No
DataSpace4Health	Healthcare	Yes	No	Low	No

⁹ <https://catena-x.net/>

¹⁰ <https://www.pontus-x.eu/>

¹¹ <https://flex4res.eu/>

Omega-X	Energy	Yes	No	No	No
Catena-X	Automotive Industry	Yes	No	No	No
Pontus-X	Cross-Domain	Yes	Indirect	Moderate	Low
Flex4Res	Manufacturing Industry	Yes	Yes	Low	No
6G-DALI	Telecom-munications	Yes	Yes	High	Yes

2.3 Alignment with other SNS and HE projects

6G-DALI is strategically aligned with key European 6G initiatives including SUNRISE-6G¹², 6G-BRICKS¹³, 6G-SANDBOX¹⁴ and ADROIT6G¹⁵, building on their foundational work in infrastructure, experimentation and network automation. While these projects focus on developing advanced testbeds, open interfaces and orchestration frameworks, 6G-DALI introduces a complementary layer centred on the lifecycle management of AI and data-driven intelligence. By integrating data space, digital twins and MLOps & DataOps capabilities, 6G-DALI enhances the usability and intelligence of 6G experimentation platforms, enabling scalable, trustworthy and AI-native experimentation for future networks. Moreover, 6G-DALI will adopt the SAT (Social Acceptance of Technology) methodological framework extended in 6G4Society¹⁶ to identify societal impacts and key values, as well as the ETHAI (Ethics of AI) methodological framework defined in the MES-CoBraD¹⁷ project that supports the development and assessment of AI technology with respect to ethics and regulatory principles.

2.3.1 SUNRISE-6G

The primary objective of SUNRISE-6G is to establish a sustainable federation of research infrastructures to significantly scale up experimentation within the 6G domain. The project's vision is built upon the "network of networks" concept, aiming to integrate both public and private infrastructures under a massively scalable, internet-like architecture. To achieve this, SUNRISE-6G seeks to create a pan-European facility that supports converged workflows and tools for experimentation. The project's execution is structured around four key pillars: the implementation of new 6G enablers, the development of a 3GPP-compliant Federation solution, the establishment of a Federated AI plane (encompassing AI as a Service and MLOps paradigms), and the creation of a commonly adopted Experimentation Plane. SUNRISE-6G will also offer a unified catalogue of 6G enablers, making them publicly accessible to experimenters and facilitating the onboarding of cross-domain vertical applications. Notably, the project leverages existing testbeds from the SNS Phase 2 projects, building upon prior investments and advancements.

Alignment with 6G-DALI: SUNRISE-6G's emphasis on a federated 6G experimentation ecosystem aligns closely with 6G-DALI's vision of an AI-driven experimentation framework. Both projects stress common workflows and on-demand experimentation. For example, SUNRISE-6G's AI plane and 6G-DALI's MLOps enable on-demand AI testing in 6G networks. However, SUNRISE-6G is primarily infrastructure-oriented – it builds the network-of-networks – whereas 6G-DALI focuses on data and AI support. Specifically, 6G-DALI provides the data pipelines, 6G Data Space and digital twin environments needed to train and validate AI models on those federated testbeds. In other words, SUNRISE-6G supplies the "fabric" of interconnected 6G resources, and 6G-DALI supplies the data management and AI orchestration layer that exploits that fabric for intelligent automation.

2.3.2 6G-BRICKS

¹² <https://sunrise6g.eu/>

¹³ <https://6g-bricks.eu/>

¹⁴ <https://6g-sandbox.eu/>

¹⁵ <https://adroit6g.eu/>

¹⁶ <https://6g4society.eu/>

¹⁷ <https://www.mes-cobrad.eu/>

The 6G-BRICKS project is dedicated to building reusable testbed infrastructures aimed at validating cloud-to-device breakthrough technologies that are envisioned for 6G networks. A core objective is to deliver a modular end-to-end 6G experimentation platform in Europe. The project places a strong emphasis on integrating breakthrough 6G technologies, such as cell-free networking, distributed processing, and Reconfigurable Intelligent Surfaces (RIS). Adopting principles of modularity and softwarisation, 6G-BRICKS aims to structure its architecture around the concept of "LEGO Bricks," delivering self-contained testbed nodes that can be easily reused across different testbed infrastructures. The project also aims to integrate with O-RAN compliant interfaces, ensuring future-proofing and interoperability of its outcomes. Furthermore, 6G-BRICKS explores novel unified control paradigms based on Explainable AI and Machine Reasoning. The project's scope includes validation and showcasing of advanced use cases in areas such as the Metaverse, holographic communication, and digital twinning.

Alignment with 6G-DALI: 6G-BRICKS is heavily focused on physical-layer innovations and modular infrastructure, whereas 6G-DALI is focused on AI/data operations. Nonetheless, both support advanced experimentation in 6G networks. For example, 6G-BRICKS provides rich data from novel radio environments (cell-free and RIS-empowered links) and exposes low-level control via xApps, while 6G-DALI offers the data management and AI training environment to exploit that data. In practice, an experiment running on the 6G-BRICKS platform (e.g. tuning a cell-free RAN slice via an xApp) could feed data back to 6G-DALI's MLOps pipeline for model training and drift monitoring. In this sense, 6G-DALI complements 6G-BRICKS by ingesting the platform's measurements into its DataOps pillar and using its digital twin to simulate large-scale scenarios.

2.3.3 6G-SANDBOX

The 6G-SANDBOX project's purpose is to support the evolution of network architectures and technologies towards 6G through the creation of an intelligent, secured, and twinning-enabled open experimentation facility. The project introduces the concept of "Trial Networks," which are fully configurable, manageable, and controllable end-to-end networks composed of both digital and physical nodes. 6G-SANDBOX aims to provide a complete and modular facility for the European experimentation ecosystem, incorporating infrastructures distributed across multiple European locations, including Athens, Berlin, Malaga, and Oulu. The facility will offer automated experimentation capabilities through a comprehensive and extensible toolbox. A key focus of the project is the validation of a wide range of promising 6G technologies and research advancements. To achieve its goals, 6G-SANDBOX has defined specific objectives related to the architecture of the experimentation facility, the development and integration of an ORAN-driven physical connectivity infrastructure, the creation of a resource management framework, and the establishment of an automated and API-based experimentation lifecycle framework.

Alignment with 6G-DALI: 6G-SANDBOX and 6G-DALI share a focus on automation and open experimentation. For example, SANDBOX's "infrastructure-as-code" and automated experiment APIs complement 6G-DALI's MLOps workflows: both aim to make 6G trials repeatable and scalable. Both projects also leverage virtualization – SANDBOX through ORAN/vRAN and 6G-DALI by orchestrating AI components. However, 6G-SANDBOX primarily addresses network and resource orchestration, whereas 6G-DALI adds a higher layer of data intelligence. In particular, 6G-DALI supplies the data lakes, analytics engines and digital twin services that SANDBOX's automated framework requires for AI-driven optimization. Conversely, 6G-SANDBOX could supply 6G-DALI with live or emulated network data (via its O-RAN testbeds) that feed into 6G-DALI's model training.

2.3.4 ADROIT6G

The ADROIT6G project proposes a set of innovations to establish a 6G network architecture designed to meet the specific requirements of next-generation applications, while achieving the ambitious KPIs defined for 6G networks. In particular, the key advancements identified by ADROIT6G as crucial for 6G network evolution include: i) leveraging AI/ML-driven optimizations throughout the network by

exploring solutions in the distributed AI space to enhance performance and enable automation; ii) transitioning to a fully cloud-native network software architecture, deployable across diverse edge-cloud platforms with built-in security embedded within the network's user layer; and iii) enabling software-driven, zero-touch operations, with the ultimate goal of fully automating all aspects of network functionality and service delivery through AI. In particular, ADROIT6G designs and implements a fully distributed AI/ML Framework for CrowdSourcing AI, which provides the AI mechanisms to assist management and orchestration components in their AI-based decisions. This framework implements the innovative concept of crowdsourcing AI, which reinterprets the concept of ML-as-a-Service (MLaaS) with the aim of streamlining the development, validation, refinement, and usage of models across technological domains, facilitating collaborative training, and the execution of inference tasks and AI services. Central to this framework is the introduction of AI/ML orchestration, which coordinates ML model training and inference, benefitting from the diversity and number of computing entities (devices, network nodes) to support the deployment of AI/ML services across multiple technological domains, facilitating the efficient sharing of ML models.

Alignment with 6G-DALI: ADROIT6G and 6G-DALI share a focus on automation of AI/ML tasks and services through innovative MLOps functionalities. In particular, both projects aim at coordinating and streamlining the management of AI/ML models on top of cloud-native distributed infrastructures that span across the edge/cloud continuum. However, while ADROIT6G focuses on a limited set of basic centralized MLOps functionalities (for coordinating training, deployment and inference tasks), 6G-DALI plans to go well beyond by providing a wide set of additional innovative distributed and cooperative MLOps functionalities and services for AI/ML models monitoring and drift detection, trustworthiness assessment, hyperparameter optimization. More importantly, it will follow a technology agnostic approach based on a meta-orchestration solution that can integrate heterogeneous MLOps tools and stack.

2.3.5 6G4SOCIETY

The 6G4Society project aims to reconcile the dual needs of advancing 6G technology's performance while embedding crucial societal and sustainable values. This initiative will foster collaboration among key Smart Networks and Services (SNS) ecosystem stakeholders, alongside civil society, regulators, policymakers, media, and the public. The goal is to develop and disseminate clear information about 6G's anticipated impacts. To achieve this, 6G4Society will leverage methodologies from ethics, law, and social sciences and humanities (SSH). This approach will facilitate an understanding of the socio-cultural dimensions of 6G, establishing a framework for responsible research and innovation (RRI). Ultimately, the project will deliver practical methods, models, guidelines, policy options, and operational recommendations for developing sustainable and socially acceptable 6G technologies and applications. Key outputs, including a Social Acceptance of 6G Technology (SAT), a Key Sustainability Indicators (KSI) framework, and policy and operational briefs, will be validated through extensive public and user engagement. This validation process will involve all 6G-SNS JU projects and the 6G-IA Working Groups, ensuring broad consensus and applicability. This comprehensive strategy seeks to ensure that 6G development is not only technologically superior but also ethically sound and socially beneficial.

Alignment with 6G-DALI: The 6G-DALI considers the SAT methodological framework for guiding the development of the Proof-of-Concepts (PoCs) for ethics, societal and regulatory dimensions. This involves identifying potential societal challenges and concerns and key value indicators (KVIs). This approach enables a continuous, iterative assessment process, fostering a better understanding of the technology and its applications.

3 Business Analysis

This section describes thoroughly the monetisation opportunities of 6G-DALI. It highlights in detail the reasons for the so-called “Data Scarcity” of telecom data sets and focuses on the strategic opportunity that this creates for the project. Additionally, it defines all the relevant stakeholders, and the value proposition 6G-DALI creates for them. A comprehensive Business Model Canvas (BMC) has been developed to serve as a complete representation of the 6G-DALI business model.

3.1 Market Analysis and Opportunities

This section investigates the Market Analysis potential for Data Spaces, the Telecommunications Industry and AlaaS in the context of 6G-DALI. Data scarcity in the 6G domain is thoroughly analysed, as well as the strategic opportunities that it creates for 6G-DALI.

3.1.1 Market Analysis for Data Spaces

Up to 2030, the rapid development of 5G/6G, AI, big data, and cloud computing give us the potential to produce yottabytes of data every year, as innovative storage technologies usher in a new era of civilization [3]. Additionally, it is anticipated that the penetration of AI into everyday life combined with quantum computing, will further amplify this trend. The term “data economy” is introduced covering all associated processes enabled by digital technologies such as data generation, collection, storage, processing, distribution, analysis, delivery and exploitation.

Data itself is not just an enabler of economic activity but nowadays is a target of government policy and corporate intention. In discussions about the value of data, it is usually compared to other traditional resources such as gas, oil or water. Each analogy focuses on certain aspects of data’s significance; however, none fully covers its nature. There are two prime features of data that differentiate them from other goods [4]: (i) data is **non-rival**, it is not exhausted by the consumers, but it can be re-used and recombined to create positive spillover effects, supporting the case for reducing restrictions on data sharing and (ii) data is more valuable when in **aggregate** i.e. leveraged through processes like analytics and machine learning, rather than when considered in isolation.

From an innovation point of view, access to data within and across ecosystems, brings limitless opportunities for groundbreaking applications such as: (i) the collection of data from moving vehicles towards the development of safe autonomous driving patterns, (ii) the energy consumption pattern from smart meters to balance power supply and (iii) the analysis of millions of molecules to find new treatments faster in the health sector.

Based on the above, there are critical challenges associated towards an entirely open and free-flowing data ecosystem, due to the risk of privacy breaches, cyber security threats or data overload. To fully harness data’s potential, Europe must not only build the **technical capacity** for data sharing and AI innovation but also develop the **governance structures** that protect citizens’ rights, ensure security, and promote ethical use.

Success lies in a balanced model that enables **access, accountability, and agility** across the entire data ecosystem.

The global big data and analytics industry is expected to experience significant growth in the next few years. It is predicted to grow at a CAGR of 14.9% between 2024 and 2032 and reach \$1.088 trillion by 2032 [5]. At this stage, only abstract estimations can be made regarding the potential share that telecom data could represent, according to which, 4% - 6% of the value could be attributed to the telecom data. This is where 6G-DALI steps in. As it will be thoroughly examined in Section 3.1.5, this creates a massive opportunity to establish a strong position within this market segment.

3.1.2 Market Analysis for 6G Networks and Telecom Data

The telecommunications industry is undergoing a major transformation driven by the convergence of AI, cloud-native technologies, and the emergence of next-generation wireless standards. As 6G moves

from vision to early-stage research and prototyping, the global market is beginning to organize around the foundational capabilities that will underpin future digital infrastructure. These include AI-native network functions, real-time data intelligence, extreme automation, and trusted data sharing mechanisms. Within this landscape, 6G-DALI addresses a unique intersection between AI, telecommunications, and data infrastructure by offering an integrated MLOps/DataOps framework and a 6G-aligned dataspace. This responds directly to current market gaps and emerging strategic needs.

According to a report by Markets and Markets [6], the global 6G market is expected to grow from USD 3.96 billion in 2030 to USD 68.69 billion by 2035. Among the most relevant drivers are the demand for intelligent, automated network operations and the rise of data-centric vertical applications such as remote surgery, industrial automation, autonomous mobility, and immersive digital experiences. These use cases place stringent latency, reliability, security, and adaptability requirements—capabilities that cannot be delivered without native AI integration and comprehensive access to high-quality, real-world data.

Regarding the monetisation of 6G data, it is considered to represent a significant economic opportunity, with the global telco data monetisation market projected to grow from USD 2.52 billion in 2023 to USD 7.8 billion by 2031 [7]. 6G-DALI supports this vision by creating a unified dataspace where stakeholders, including network operators, vertical service providers, and AI developers, can contribute, share, and monetize their data. This dataspace will support multiple revenue streams, including pay-per-use data access, data licensing, and AI-driven analytics services, enabling 6G stakeholders to extract maximum value from their data investments.

Key trends shaping the 6G ecosystem further reinforce the relevance of 6G-DALI. First, the push toward sovereign and interoperable data ecosystems—promoted by initiatives like Gaia-X and the European Data Strategy—redefines how data is accessed, exchanged, and monetized. 6G-DALI contributes to this transformation by delivering a 6G Data Space that supports data sharing among telecom operators, AI developers, and vertical service providers under clear governance rules. Second, the increasing adoption of MLOps and DataOps in industrial and telecom environments signals a shift toward continuous AI deployment pipelines, where retraining, monitoring, and performance validation must occur in near-real time. The platform developed within 6G-DALI directly supports this operational model, reducing lifecycle costs and accelerating innovation.

Moreover, standardization bodies (e.g., 3GPP¹⁸ and ITU-T¹⁹), research alliances (e.g., 6G-IA²⁰ and NGMN Alliance²¹) and the Smart Networks and Services Joint Undertaking (SNS JU), have identified AI integration, testbed federation, and digital twins as key building blocks for future 6G systems. 6G-DALI's alignment with these trends positions the project to have significant technical and commercial influence. It can reduce vendor lock-in, enhance interoperability, and offer a neutral, European-driven platform for stakeholders seeking to develop, validate, and deploy AI-driven 6G applications across sectors.

6G-DALI responds to a confluence of urgent market needs and policy priorities by delivering the technological and governance infrastructure needed for a thriving AI-native 6G ecosystem. Its contributions are timely, strategic, and highly relevant to emerging trends that will define the competitiveness and sovereignty of Europe's digital future.

3.1.3 Market Analysis and Potential for AI-as-a-Service (AIaaS) in 6G-DALI

As the foundation for AI-native networks is established, AI-as-a-Service (AIaaS) is rapidly gaining traction as a commercial and operational model with strong relevance to the 6G-DALI framework. AIaaS enables the delivery of modular, cloud-based AI capabilities – including model training, inference, analytics, drift detection, and orchestration – as on-demand services. This model significantly lowers

¹⁸ <https://www.3gpp.org/>

¹⁹ <https://www.itu.int/en/ITU-T/Pages/default.aspx>

²⁰ <https://6g-ia.eu/>

²¹ <https://www.ngmn.org/>

the barriers to AI adoption by decoupling AI functionality from proprietary infrastructure and making it available via standardized interfaces, which aligns directly with 6G-DALI's MLOps and DataOps architecture. The project's integrated platform offers not only the technological substrate for AIaaS delivery, but also the governance, lifecycle management, and federated data access mechanisms required to make such services reliable, scalable, and compliant with European data regulations.

According to Markets and Markets [8], the global Artificial Intelligence-as-a-Service (AIaaS) market is projected to grow from USD 9.3 billion in 2023 to USD 43.3 billion by 2028, at a compound annual growth rate (CAGR) of 36.2% over the forecast period. This accelerated growth reflects a strong and growing demand across industries for scalable, cloud-based AI services that reduce infrastructure costs and simplify deployment. For telecom operators, vertical service providers, and developers working within the 6G-DALI ecosystem, this trend underscores the strategic relevance of providing and consuming AI capabilities in a modular, service-oriented format. Within 6G-DALI's architecture, AIaaS becomes a practical and commercially viable model enabled by its Digital Twin testbed, MLOps pipeline, and federated 6G Data Space. These components allow stakeholders to access high-performance AI functions – such as model training, inference, and orchestration – via secure, standards-based APIs. This not only supports operational efficiency but also unlocks new revenue models around AI data-set-generation-as-a-service, data set subscriptions, and shared model marketplaces.

The AIaaS market growth is attributed to several factors:

- **Proliferation of Big Data and Real-Time Analytics:** Enterprises are generating vast amounts of data, necessitating advanced analytics tools. AIaaS provides the infrastructure to process and analyse this data efficiently.
- **Adoption Across Diverse Industries:** Sectors such as healthcare, finance, retail, and manufacturing are increasingly leveraging AIaaS for applications like predictive analytics, customer service automation, and operational efficiency.
- **Emergence of Low Code/No Code Platforms:** The development of user-friendly AI platforms allows organizations without extensive technical expertise to implement AI solutions, broadening the market reach.
- **Integration with Emerging Technologies:** AIaaS is being combined with technologies like edge computing and the Internet of Things (IoT), enabling real-time data processing and decision-making at the source.

In the context of the 6G-DALI, these trends underscore the strategic importance of integrating AIaaS into next-generation telecommunications infrastructure. AIaaS allows users to deploy and consume machine learning functions, inference models, and data pipelines through cloud-native platforms, with seamless integration into vertical applications and network environments. This consumption model aligns naturally with the needs of the telecommunications industry, particularly in the context of 6G, where continuous intelligence and responsiveness are essential features of future networks.

Current market dynamics indicate strong momentum behind AIaaS adoption, fuelled by the convergence of multiple drivers: increasing demand for real-time analytics, pressure for operational efficiency in data-intensive environments, and the need for agility in AI deployment cycles. The shift toward “as-a-service” paradigms is not merely a technological preference but a business necessity, enabling organizations to shift capital expenditure (CAPEX) to operational expenditure (OPEX), optimize resource utilization, and accelerate time-to-value for AI initiatives. In highly dynamic and regulated sectors such as telecom, this model provides both economic and compliance advantages.

Industry analysts highlight that enterprise demand for AIaaS is moving beyond generic services toward domain-specific, trusted, and interoperable solutions [9]. In this regard, 6G-DALI would offer a differentiated market position. The platform enables the delivery of AI functionalities through secure APIs that are directly embedded into the operational logic of 6G systems—offering capabilities such as traffic prediction, anomaly detection, orchestration intelligence, and network slicing optimization. These services can be provided on-demand to vertical actors, service providers, and application

developers, creating a modular ecosystem where intelligence is no longer static or centralized, but distributable and customizable based on context.

Moreover, the integration of AIaaS into a shareable data and model environment, such as the 6G Data Space developed in the project, adds a governance and trust layer that is increasingly demanded by European and international customers. This is especially relevant for cross-border AIaaS offerings that must comply with data sovereignty, privacy, and ethical AI requirements. By aligning its design with frameworks such as Gaia-X and the European AI Act, 6G-DALI ensures that its AIaaS architecture can be positioned as a compliant and competitive offering in both public and commercial procurement environments.

As organizations transition from AI experimentation to production-scale deployment, the need for integrated pipelines that combine data access, model execution, monitoring, and retraining becomes more pronounced. 6G-DALI can address this lifecycle by embedding MLOps workflows within the AIaaS model, enabling intelligent services to adapt over time to network conditions, policy changes, or new vertical requirements. This end-to-end automation also opens the door to new business models, including AIaaS subscriptions and dynamic AI marketplaces where stakeholders can exchange models and training assets. The growing maturity of the AIaaS market presents a unique opportunity for 6G-DALI to serve as both a technological foundation and a commercial enabler. Its platform-centric approach, reinforced by strong alignment with EU strategic goals for AI and data sovereignty, allows it to support a wide spectrum of actors – from operators to SMEs - in developing scalable, adaptive, and trusted AI services for the 6G era.

3.1.4 Data Scarcity in the Telecom domain

Focusing on the Telecom domain, related data are notably scarce. In the context of data spaces and in initiatives like the Common European Data Spaces (CEDS)²², this scarcity is attributed to regulatory, technical and organizational challenges. More specifically:

- **Regulatory Constraints** - strict constraints such as those imposed by the General Data Protection Regulation (GDPR) require anonymization and encryption mechanisms, which make it complex for telecom operators to share any data without risk of non-compliance.
- **Privacy Concerns** - telecom data, encompassing subscribers' call records, location information and general behaviour of the subscribers is highly sensitive. Based on that, operators cannot risk the severe and far-reaching consequences of a data breach involving subscriber information, as such an incident could result in catastrophic damage to their brand reputation, erode customer trust, and invite regulatory penalties. Protecting customer data is not only a legal obligation but a critical component of maintaining competitive credibility and long-term viability in an increasingly data-sensitive market.
- **Data Silos** – legacy systems still exist within telecom organisations. These systems are not designed for smooth data integration with other systems. Data silos cause lack of interoperability which ultimately hinder the exchange of data. In conventional data management environments, information is often stored in isolated silos, each functioning independently with minimal interaction or integration. Introducing new data sources or formats into these silos frequently results in data duplication, quality inconsistencies, and integration difficulties. Moreover, data silos hinder effective data sharing across broader networks of partners and ecosystem stakeholders, significantly restricting the potential for data reuse and collaboration [11].
- **Lack of Standardization** – there is a lack of standardized approaches to data sharing between telecom operators. This absence of data uniformity leads to inconsistencies in data quality across different regions and providers. Each telecom provider typically maintains its own proprietary systems, data schemas, and internal governance processes. As a result, there is a lack of uniformity in how data is structured, accessed, and exchanged across the sector. This **data fragmentation** leads to inconsistencies in data quality. The implications are particularly significant in the context

²² <https://digital-strategy.ec.europa.eu/en/policies/data-spaces>

of global telecommunications, where services and infrastructures increasingly span multiple countries and networks. Without consistent standards, interoperability is limited.

- **Lack of Incentives** - Telecom operators lack incentives to participate in frameworks for data sharing, exchange and reusability, such as data spaces. A clear business case is missing at this stage to convince organisations to ‘invest’ time and effort into making their data available for external use. It has to be considered that there are substantial costs involved for data extraction, processing and ensuring compliance with standards and a business model should address those issues and describe the incentives.

3.1.5 Strategic Opportunity in Addressing the 6G Data Gap

The transition to 6G networks introduces unprecedented technical capabilities and a significant and under-addressed challenge: the lack of high-quality, large-scale, and representative data sets tailored to the unique requirements of AI-native 6G environments. This scarcity of data, unlike any faced in prior mobile network generations, presents both a critical bottleneck and a transformative market opportunity. Without a concerted effort to close this gap, the ambitions for intelligent, adaptive, and real-time 6G systems risk being delayed or derailed due to poorly trained AI models and limited cross-sector experimentation.

6G networks are expected to integrate ultra-low-latency communications, real-time AI processing, semantic and goal-oriented networking, and massive-scale connectivity across devices, users, and applications. These capabilities demand AI systems trained not only on high-fidelity data but also on data that captures the complexity and heterogeneity of future network conditions. However, the current state of AI model development for mobile networks remains heavily dependent on simulated, outdated, or contextually narrow data sets—nonsufficient to support the intelligence required for 6G.

This data gap stems from several interdependent factors:

- **Novel Use Cases and Disruptive Network Paradigms:** 6G envisions use cases that go far beyond the capabilities and contexts of 5G. Applications such as immersive extended reality (XR), high-fidelity holographic communication, real-time digital twins, human-centric services, and pervasive intelligence at the edge introduce data types, traffic patterns, and behavioural dynamics that are fundamentally new. These use cases generate multi-modal, context-sensitive data, ranging from visual and spatial inputs to real-time sensor streams and user intent data, that currently do not exist at scale. Training AI models to function effectively in these environments requires extensive and context-rich data sets that reflect the nuances of such novel scenarios.
- **Fragmentation and Complexity in Data Collection:** 6G networks are inherently heterogeneous. They will operate across diverse geographies and deployment conditions, from ultra-dense urban cores to rural industrial corridors. They will incorporate multiple tiers of intelligence—from edge and fog to centralized cloud—and engage a wide array of data-generating sources, including autonomous vehicles, IoT devices, AR/VR systems, and next-generation base stations. Capturing, cleaning, and harmonizing such diverse data in a scalable, privacy-aware, and regulation-compliant manner presents significant technical, operational, and economic hurdles. Moreover, traditional data collection pipelines, which rely on manual integration and offline pre-processing, are poorly suited to the dynamic and distributed nature of 6G environments.
- **Limited Access to Operationally Realistic Testbeds:** The maturity of AI systems depends critically on the quality and realism of their training environments. However, access to real-world or high-fidelity testbeds for 6G experimentation is currently limited, both in number and scope. Most available environments are either too small in scale, too constrained in terms of scenario diversity, or too disconnected from operational networks to provide meaningful data for AI model development. As a result, developers and researchers are often forced to rely on outdated data sets, leading to poor generalizability and performance drift when models are deployed in production.

3.1.6 6G-DALI's Strategic Response

6G-DALI is uniquely positioned to transform this challenge into a long-term advantage through its integrated DataOps architecture and comprehensive 6G Data Space. Recognizing that data is not simply a by-product but a core enabler of AI-native networking, the project introduces a structured, secure, and sovereign approach to data lifecycle management that directly addresses the bottlenecks identified above. 6G-DALI aims to address this lack of data and to create strategic value by establishing a dedicated 6G Data Space, as part of its broader DataOps pillar. This Data Space will provide the foundational data infrastructure for 6G by:

- **Increasing Data Availability:** The 6G-DALI project addresses the critical challenge of data accessibility in next-generation telecommunications through the establishment of a comprehensive data space that dramatically enhances the discoverability of 6G datasets. By implementing a data infrastructure with standardized cataloging and metadata management systems, 6G-DALI transforms previously siloed and fragmented 6G research data into discoverable, accessible resources for the broader telecommunications community. The data space architecture incorporates sophisticated search and discovery mechanisms that enable researchers, developers, and industry stakeholders to efficiently locate relevant datasets across multiple testbeds and research initiatives. This enhanced discoverability is achieved through comprehensive metadata schemas that describe dataset characteristics, experimental conditions, and usage parameters, coupled with semantic vocabularies that ensure consistent interpretation across different research contexts. The strategic implementation of this data availability framework not only accelerates research and development cycles by reducing data acquisition barriers, but also maximizes the value extraction from existing 6G experimental investments by enabling cross-project data utilization and collaborative analysis. Through this systematic approach to data discoverability, 6G-DALI creates a multiplier effect that amplifies the impact of individual research efforts while fostering a more collaborative and efficient 6G research ecosystem.
- **Facilitating Testbed Integration:** The 6G-DALI Data Space architecture is designed with extensibility and scalability at its core, enabling seamless integration of additional testbeds into the telecommunications research ecosystem. Through standardized dataspace connectors and well-defined integration protocols, new testbeds can be onboarded efficiently without requiring extensive custom development or architectural modifications to the existing infrastructure. The modular design of the ingestion layer ensures that diverse testbed environments, regardless of their underlying technologies or data formats, can be accommodated through configurable connector components that handle data transformation, protocol translation, and metadata harmonization. This streamlined integration approach significantly reduces the technical barriers and time-to-market for testbed participation, encouraging broader community engagement and expanding the diversity of available research data. The standardized metadata schemas and semantic vocabularies provide a consistent framework that new testbeds can adopt, ensuring immediate interoperability with existing datasets and analytical tools upon integration. By lowering integration complexity and providing clear technical specifications, 6G-DALI creates an attractive proposition for research institutions and industry partners to contribute their experimental data, thereby continuously enriching the collective knowledge base and strengthening the overall value proposition of the 6G research ecosystem.
- **Accelerating AI Development with High-Quality Data Pipelines:** By implementing full-lifecycle **DataOps practices**, 6G-DALI ensures that data used for AI model training is of high quality, versioned, traceable, and compliant. These practices include automated data validation, anomaly detection, metadata tagging, and lineage tracking, which enhance model robustness, reduce retraining costs, and shorten the time from development to deployment. For example, AI models trained with such rigorously curated data are more likely to generalize across network domains and user populations, reducing the risk of bias, error propagation, and performance drift.
- **Supporting Sovereign Data Sharing:** Aligned with the principles of Gaia-X and the European Data Strategy, the 6G-DALI Data Space promotes trust-based data sharing among stakeholders. By

logically centralizing data storage, we ensure governance and security via a platform that facilitates data sharing and discoverability via standardized APIs and smart contracts. This promotes collaborative innovation while respecting data sovereignty, ownership rights, and regulatory frameworks such as the GDPR and the upcoming AI Act.

- **Enhancing Cross-Industry Innovation:** The unified and standardized nature of the 6G-DALI Data Space facilitates cross-industry collaboration. Vertical domains such as healthcare, automotive, logistics, and smart cities can contribute to and benefit from domain-specific data sets. This mutual reinforcement enhances the development of specialized AI models tailored to each sector's operational and regulatory nuances, unlocking new applications and services while promoting interoperability and reuse. It also enables public-private cooperation in co-developing data sets for societal missions such as crisis response, green mobility, or digital health.
- **Enhancing Data Availability through Digital Twins:** Through its Digital Twin Testbed, 6G-DALI supports the creation of synthetic data sets that realistically simulate 6G traffic and operational conditions across a range of use cases and environments. These synthetic data sets, generated with contextual fidelity, can serve as stand-ins for real-world data, enabling early-stage AI development without the need for large-scale physical deployments. Additionally, the project integrates mechanisms for capturing real operational data in testbed environments, ensuring that models trained on synthetic data can be fine-tuned and validated with real-world inputs. This dual approach ensures both agility and realism in AI development.

By addressing the lack of 6G-ready data through a combination of synthetic generation, real-world testbed integration, sovereign data sharing, and industrial-grade DataOps, 6G-DALI is not merely responding to a challenge—it is establishing a foundational enabler for the entire 6G AI ecosystem. The project's approach empowers all stakeholders to innovate faster, with greater confidence, and in alignment with European values of privacy, trust, and inclusiveness. 6G-DALI is a critical enabler for the AI-driven 6G ecosystem, unlocking new value for all stakeholders and ensuring a competitive edge in the rapidly evolving telecom landscape.

3.2 6G-DALI's Stakeholders and Impact

The business plan of 6G-DALI outlines how the project's core assets — namely, telecom-grade data sets, advanced MLOps/DataOps capabilities, and integrated experimentation infrastructure — can serve as enablers for long-term value creation within the European 6G ecosystem. 6G-DALI focuses on establishing the technical, operational, and governance conditions necessary for stakeholders to access, generate, and reuse high-quality 6G data in a trusted and sovereign manner. This section defines the key stakeholder groups, identifies the potential benefits they derive from the platform, and presents a business model that reflects 6G-DALI's role as a foundational infrastructure for AI-driven 6G innovation. The approach aligns with European strategic priorities for data sharing, openness, and digital sovereignty.

3.2.1 Identification of the 6G-DALI stakeholders and their roles

6G-DALI introduces a transformative end-to-end AI framework built on MLOps and DataOps, enabling intelligent, automated management of future 6G networks. The following stakeholders are involved in or impacted by the framework either directly or indirectly.

1. **Network/Cloud Provider (NCP)** (e.g. Mobile Network Operators (MNOs), MVNOs, Telco Vendors, and Brokers.) Network Operators maintain and operate a 6G network infrastructure and offer connectivity services, across part or all segments of the network. The network service may be offered in the form of Network Slice-as-a-Service (NSaaS) or Network-as-a-Service (NaaS), e.g., public network integration with non-public network (PNI-NPN). They also require tools for service assurance, drift detection, and performance monitoring that are built and improved using available or collected 6G data sets.
2. **Data Platform as a Service Provider (DPaaSP).** Service providers offering managed platforms for hosting, sharing, and operating on data and AI assets in the cloud. In 6G-DALI, they enable the

- deployment and operation of the dataspace, ML models and datasets by providing trusted environments for data exchange, collaboration, and lifecycle management.
3. **Vertical Service Provider (VSP).** This stakeholder represents a vertical company/organization (SMEs or enterprises) that deliver domain-specific services (e.g., public safety, smart cities, eHealth). This role is also commonly referred to as tenant (5GPPP Arch. WP., 2017[10]). They act as experimenters, leveraging 6G- DALI to deploy, test, and optimize AI-enhanced services, as well as to run data experiments for generating and collecting 6G datasets. They benefit from ML predictors, demand forecasting, and seamless integration with communication and application services.
 4. **AI experts and ML developers (AEMD).** Stakeholders from academia and industry creating AI models for 6G applications. They leverage 6G-DALI's data generation and improvement capabilities (e.g., AutoML, dataset repositories, data generation), advanced MLOps tools (for model training, hyperparameter optimization, finetuning, testing, and deployment) and DataOps services to overcome data set limitations, accelerating AI solution development.
 5. **Cloud Infrastructure Service Provider (CISP).** Stakeholder managing physical or virtualized infrastructure (e.g., virtual machines (VMs) / containers, virtualised / physical infrastructure (CPUs, GPUs, etc) that includes computational, storage, networking or IoT resources) hosting AI training and inference tasks. Supports on-premises, edge, or hybrid scenarios, and enables compute/storage elasticity required by MLOps pipelines.
 6. **Framework Service Provider (FSP).** Stakeholder that offers the framework software services.
 7. **Data Provider (DP).** Entity (e.g., a research institute or telco broker) capable of leasing testbed resources (RAN, Core, Edge/Cloud) commercially as well as digital twins, that generate and provide data. Ensures SLA-based availability for the AI application development, enabling data set generation as a service.

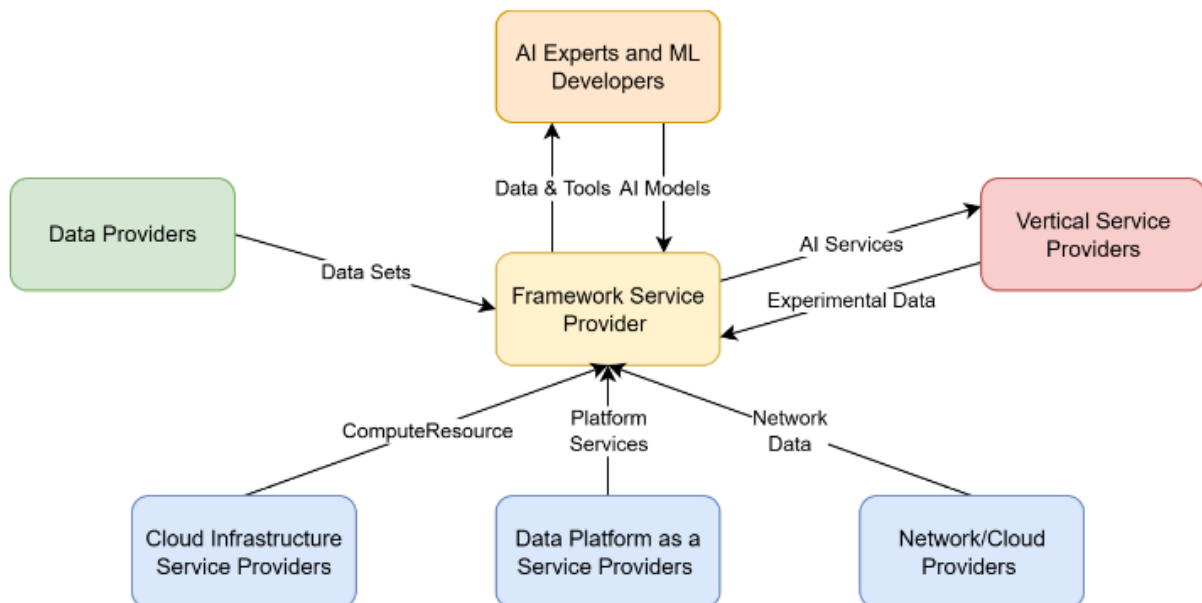


Figure 2: Relationship between 6G-DALI's stakeholders

Figure 2 depicts the relationship between 6G-DALI's aforementioned stakeholders. Each stakeholder identified, derives value from the existence, accessibility, and governance of the high-quality datasets produced within 6G-DALI. The specific ways in which these data assets are used—ranging from AI model training and validation to service orchestration and marketplace offering—are elaborated in the stakeholder-specific value propositions in Section 3.2.3.

3.2.2 6G-DALI Impact

The impact of the project is twofold—technological and socio-economic. Technologically, it provides the building blocks for a new class of intelligent, adaptive 6G networks by integrating AI workflows (MLOps)

and data lifecycle tools (DataOps) in a unified architecture. This supports creating, validating, and deploying AI models in realistic network environments, accelerating innovation and reducing time-to-market for new services. Economically, the project contributes to developing a European data economy by enabling monetization models around data exchange, synthetic data generation, and AI service orchestration. The proposed 6G Data Space, aligned with Gaia-X principles, enables trusted data set aggregation and ecosystem-level collaboration by enforcing distributed governance and clear data usage policies. While data is technically stored in a centralized manner for operational efficiency, reducing the need for data storage infrastructures deployed in the interconnected testbeds, access and control remain interconnected, ensuring transparency, interoperability, and data sovereignty. This approach guarantees that SMEs, startups, and research institutions can participate on equal footing with larger industrial players, without being locked into monolithic cloud architectures or losing ownership of their data assets.

6G-DALI is expected to enhance Europe's strategic autonomy in key technological areas. Reducing dependency on proprietary data platforms and promoting open standards strengthen the resilience of European telecom infrastructure and support the creation of sustainable, trust-based digital ecosystems. The project also contributes to key policy goals of the European Commission, including the twin green and digital transitions, by enabling energy-efficient AI operations and data-driven optimization in sectors such as mobility and health. It addresses a high-priority market demand while laying the foundation for long-term economic and technological impact. By fostering a vibrant ecosystem of stakeholders, supporting data sovereignty, and enabling AI innovation at scale, it contributes directly to Europe's competitiveness in the 6G era and positions the project as a key enabler of the future digital infrastructure.

3.2.3 6G-DALI Value Proposition to Stakeholders

6G-DALI offers a data-centric value proposition to a diverse set of stakeholders by delivering a comprehensive end-to-end AI framework that integrates MLOps and DataOps capabilities, tightly coupled with the availability, generation, and sharing of 6G-relevant data sets. Rather than targeting a new market or proposing entirely new tools/products, 6G-DALI targets the existing telecommunications infrastructure market and aims to deliver economic benefits by reducing operational expenditure through enhanced automation and efficiency. This framework addresses the critical need for high-quality data and AI automation in next-generation networks. By integrating **MLOps-driven AI experimentation** and **DataOps-enabled data lifecycle management**, the project equips each stakeholder group with concrete tools, capabilities, and efficiencies that align with their domain-specific needs. Specifically, the value proposition offered to each of the project's stakeholders includes:

1. **Network/Cloud Providers (NCP)** gain significant operational benefits from the 6G-DALI framework, particularly through its AI-driven capabilities for resource optimization, fault detection, and service assurance. By leveraging the MLOps and DataOps pipelines provided by the framework, NCPs can develop and deploy machine learning models tailored to dynamic network conditions. These models rely on continuous access to curated, high-quality data sets, both real and synthetic generated via the integrated Digital Twin testbed.
2. **Data Platform as a Service Provider (DPaaS)** can host and scale machine learning models within the secure 6G-DALI Data Space, supporting data-driven innovation and providing a differentiated, AI-ready infrastructure for 6G applications. The value of this offering stems from the availability of well-governed and telecom-specific data sets, which feed the training pipelines and enhance the AI-readiness of cloud platforms in the telecom domain.
3. **Vertical Service Providers (VSP)**, including sectors like eHealth and smart cities, can leverage the framework for deploying and optimizing AI-enhanced services, reducing time-to-market and enhancing operational reliability. By accessing curated and telecom-aligned data sets, verticals can test and adapt their services to real network conditions before deployment, ensuring higher performance and cross-domain compatibility.

4. **AI Experts and ML Developers (AEMD)** are empowered with advanced MLOps tools for model training, hyperparameter optimization, and deployment, significantly reducing the time and complexity involved in AI solution delivery for 6G. In parallel, they gain access to a structured environment with trustworthy, versioned data sets and built-in data pipelines, which dramatically reduce data wrangling efforts and enable rapid, reproducible model iteration across diverse 6G scenarios.
5. **Cloud Infrastructure Service Providers (CISP)** play a crucial role in supporting the physical infrastructure needed for training and deploying AI models, benefiting from the scalability and flexibility of the 6G-DALI framework. The growing demand for data-driven experimentation and storage places data availability at the centre of infrastructure design, with CISPs positioned to offer value-added services around secure and scalable data set handling.
6. **Framework Service Providers (FSP)** play an important role in offering the framework software services.
7. **Data Providers (DP)** can commercialize their infrastructure by offering "data set generation as a service," leveraging the 6G-DALI framework to attract AI and 6G researchers and businesses. This opportunity is powered by the data sets generated through testbed activity, which become valuable assets when shared, reused, or augmented within the 6G-DALI Data Space.

3.3 Business Models

This section presents the Gaia-X Business Model, and it is explained how 6G-DALI integrates in business terms with the ecosystem. The Section is concluded with a Business Model Canvas (BMC) which provides a consolidated representation of the proposed Business Model.

3.3.1 Gaia-X Business Model

The Gaia-X business model is built around the principles of secure data exchange, transparency, and sovereignty. It aims to establish a federated data ecosystem where data and services can be shared seamlessly, but always under strict control of data origin, quality, and access rights. Monetisation with the help of Gaia-X is achieved both through offering innovative services and by leveraging high-value, authenticated data sets to potential client organizations—empowering businesses with global competitive advantage. Table 3²³ describes all the potential types of business model which could provide benefits to stakeholders via Gaia-X principles.

Building on this foundation, the Gaia-X business model directly informs and shapes the 6G-DALI business model seeking to the sustainability of 6G-DALI after the completion of the project. By adopting Gaia-X's principles, the 6G-DALI framework creates a trusted data space where diverse stakeholders—including AI service providers, network operators, and application developers—can confidently collaborate and innovate. This trusted environment supports primary and secondary monetization streams, enabling participants to commercialize not only their AI-driven services but also the valuable data assets and supporting infrastructure. The competitive edge is especially pronounced as 6G-DALI participants, like those in Gaia-X, benefit from access to authenticated, high-quality data sets. This accelerates AI development, enhances reliability and security, and enables sustainable differentiation in a rapidly evolving 6G ecosystem.

²³ <https://gaia-x-hub.de/wp-content/uploads/2023/02/Whitepaper-Gaia-X-Business-Models.pdf>

Table 3: Business Model Types enabled by Gaia-X.

	Primary Classification	Secondary Classification	Description	Monetization
Gaia-X Business Model Types	Consumer	Data	A company acquires data from different sources and offers it. The data is subsequently offered via Gaia-X after several processing and documentation steps.	The company, along with AI providers in general gains a global competitive advantage with Gaia-X, because its own models can be trained on data of the highest quality with an authenticated origin.
		Data Analysis	The consumers receive insights that enable them to better implement their current business model, optimise it or open up entirely new business opportunities.	A service provider can be found even for very specific requirements. The necessary data and inputs can also be obtained (from the provider or consumer) via Gaia-X-compatible data spaces
	Provider	Infrastructure	Companies in this sector procure infrastructure services. For example, they may be companies that offer data analytics services. They may, for example, demand storage space for their data or development platforms for the development of their analytics services.	It offers uniform standards on the basis of which providers of the most diverse size and focus can become active. solutions can be found that are tailored to the respective business model and, for example, represent an alternative to the large established cloud providers.
		Data	A company acquires data and offers it. The data can come from different sources. For example, data can be generated through the use of sensors or collected through user input. The data is then offered via Gaia-X, possibly after several processing and documentation steps	Monetization can take place in various forms, for example through subscriptions or one-off payments. This business model is also of particular interest to small and medium-sized companies (SMEs).
	Federator	Data Analysis	The companies in this category offer data analysis as a service.	They monetize their analytical capabilities, with Gaia-X opening up new customer segments for them and thus enabling the company's business model.
		Infrastructure	Providers of cloud services such as: (i) Infrastructure as a Service, (ii) Platform as a Service, or (iii) Software as a Service (Gaia-X Architecture Document, p.54).	This makes it easier to develop and implement specialized offerings and new business models beyond the current services offered by the large cloud providers
		Data Exchange (technical)	Companies in this category provide federation services, such as: (i) secure transmission of data, (ii) the cataloguing of offerings, (iii) the identification of participants, etc	Monetisation is by means of payments to the companies for the provision of these services, either by the users or the operators of a dataspace.

3.3.2 6G-DALI Business Model

The 6G-DALI aims to establish the first European 6G Data Space that aligns closely with the principles and architecture of Gaia-X. The primary goal is to create a secure data infrastructure that supports the efficient exchange, processing, and monetization of 6G data and data services. This approach is critical for the technical and economic viability of 6G, as it addresses the significant challenge of generating high-quality, representative data sets for AI training, validation, and deployment. The creation of a dedicated 6G Data Space as envisioned by the 6G-DALI project marks a pivotal milestone in enabling AI-native, data-driven networks.

From a technical perspective, the adoption of standardized data connectors, secure data exchange, and federated identity management principles lies in the heart of the 6G Data Space. These components are critical for ensuring that 6G data can be shared across different networks and organizations without compromising privacy or data sovereignty. Such an architectural decision reduces integration costs, promotes collaboration between enterprises, links data producers and data consumers, accelerates innovation, and provides a trusted foundation for AI-driven 6G services.

Economically, the 6G Data Space creates new opportunities for value creation by enabling data-driven business models. By facilitating the exchange of 6G data in a standardized, secure manner, 6G-DALI allows telecom operators, cloud providers, and vertical service providers to build, test and validate novel services over real-world data generated by 6G testbeds. These services can benefit from the DataOps and MLOps provided by 6G-DALI to speed up their development cycles and streamline their operations, thus reduce costs for data acquisition, data processing and solution development, while generating new revenue streams through data monetization. This approach also supports the broader goals of the European Digital Strategy, reinforcing the EU's digital sovereignty and competitiveness in the global technology market.

To support these opportunities in practice, 6G-DALI outlines a set of potential revenue enablement paths that emerge from its core technological capabilities. These paths reflect how the project's assets — including AI-ready data sets, testbed infrastructure, and modular MLOps/DataOps tools — can be reused or extended to support services and interactions beyond the project lifetime. While not intended as direct commercial offerings during the project, these streams represent realistic and sustainable value creation mechanisms for stakeholders across the telecom and AI ecosystems. The four key areas identified in the business model include:

Licensing for DataOps and MLOps Platform Features: The 6G-DALI architecture introduces modular and reusable components for both **DataOps** and **MLOps**, designed to manage data set quality, versioning, and model lifecycle automation. These components are not rigidly bundled as a product suite but are offered as open or semi-open modules that can be selectively adopted by stakeholders looking to operationalize AI in telecom and vertical contexts. Licensing opportunities may emerge for specific capabilities, such as **automated data pipelines, integrated drift detection, metadata management, or validation workflows**, particularly when these tools are adapted for use beyond the scope of the project. This licensing model does not imply commercialization within the project timeline but reflects a **post-project sustainability path**, especially for academic spin-offs, industrial collaborations, or open-source-driven service models. By maintaining modularity, the platform allows selective licensing of core functionalities or support services, depending on stakeholder maturity. This is consistent with the project's goal of **lowering the adoption barrier for AI integration in 6G** and enabling trusted data/ML operations through flexible and reusable software artifacts.

Subscription-Based Access to Data sets, Tools, and Testbeds: One of 6G-DALI's key contributions is the exposure of curated data sets — both **real (from testbed operations)** and **synthetic (from Digital Twins)** — in a structured and annotated form. Combined with integrated MLOps and DataOps tools, this data becomes **AI-ready** and suitable for a variety of tasks, including training, benchmarking, or transfer learning. A subscription-based access model can enable different user tiers (e.g. researchers, SMEs, vertical providers) to access data sets and tools under usage or attribution terms, especially in

post-project federated or cloud-hosted environments. The same model applies to the **controlled access of testbeds** for remote experimentation, where subscribers can deploy models, observe behaviour, and validate services across realistic 6G network conditions. Subscription does not imply traditional paywalling but can be adapted to **community access frameworks**, quota-based models, or institutional support programs — **in line with Gaia-X-aligned, non-discriminatory access principles**. This reinforces sustainability while preserving openness and reproducibility, key pillars of the project's technical and ethical design.

Monetization of Real World and Synthetic Data and AI Data-Set-Generation-as-a-Service: Through its real world and Digital Twin scenario generators, 6G-DALI can produce data sets that replicate telecom behaviour under configurable conditions — traffic surges, mobility events, radio channel anomalies, etc. These data sets, validated, offer high utility for training and testing AI models in domains even where real data may be scarce or privacy-restricted. Monetization here refers not to commercial sales but to structured, compliant reuse models, where third parties (e.g. startups, verticals) may contribute to or benefit from data pools. In parallel, 6G-DALI enables AI data-set-generation-as-a-service, where external users can execute workflows — data set generation, data cleaning, data augmentation — on data from integrated testbeds with data set support and orchestration logic. This model is particularly valuable for researchers and SMEs lacking access to physical infrastructure or data curation capacity. It allows the project infrastructure to act as a trusted data layer post-project, potentially coordinated through innovation hubs, competence centres, or public-private labs aligned with the EU's digital objectives.

Consultation and Training for Integration, Maintenance and Deployment: Given the technical complexity of integrating AI with telecom systems, **training and onboarding support** becomes a natural service path emerging from 6G-DALI. This includes documentation, tutorials, and guided sessions on how to use DataOps and MLOps components, interact with the testbed APIs, and prepare data sets for AI workflows. Such activities support broader capacity-building goals and could be offered via workshops, online platforms, or technical assistance programs beyond the project. Furthermore, as verticals and SMEs seek to adopt 6G-DALI components into their own environments, **consultation services** — for example, integration with local data platforms, adaptation of synthetic data workflows, or model validation — may become relevant. These services don't imply commercial consultancy but align with sustainability mechanisms, such as **participation in national research infrastructures, competence centres**, or cross-project technical alliances. They support 6G-DALI's long-term goal to **bridge the gap between experimental research and operational deployment** across sectors.

One of the most promising aspects of the 6G-DALI business model is the offering of 6G data through a standardized interface. The global market for telecom data monetization is expected to grow significantly in the coming years, driven by the increasing demand for real-time, AI-powered services. 6G-DALI positions itself to capture this opportunity by enabling stakeholders to benefit from their data in several ways:

- **Data Marketplaces:** 6G-DALI can facilitate the creation of compliant data marketplaces, where stakeholders can offer their 6G data to third parties for purposes such as AI training, network optimization, and predictive analytics.
- **Data-as-a-Service (DaaS):** Stakeholders can offer their data as a service, charging for real-time data streams, historical data, or analytics insights. This model aligns well with standardized principles of data sovereignty and federated data exchange.
- **Value-Added Services:** Beyond raw data sales, stakeholders can offer higher-margin, AI-driven services like predictive network management, anomaly detection, and performance optimization, significantly enhancing the monetization potential of their data assets.

This multi-layered approach to data exploitation not only generates direct benefits but also creates long-term strategic advantages by positioning 6G-DALI as a critical data infrastructure provider for the 6G ecosystem. By providing a trusted, interoperable platform for data sharing, 6G-DALI supports faster AI model development, reduced operational costs, and improved service quality. This, in turn, strengthens

the competitive position of European technology providers and helps achieve the EU's digital sovereignty goals.

3.3.2.1 Business Model Canvas

To articulate the commercial and strategic potential of the 6G-DALI project, a Business Model Canvas (BMC) has been developed as shown in Table 4. The BMC provides a structured overview of how 6G-DALI creates, delivers, and captures value within the emerging 6G ecosystem. It highlights the project's key partners, core activities, value propositions, and the various stakeholder groups it addresses.

In particular, 6G-DALI's business model reflects its dual mission: to bridge a critical technological gap in AI-native 6G networks and to support a sustainable data economy via an open, secure, and federated data-sharing infrastructure. This model emphasizes the importance of collaboration across vertical industries, research organizations, cloud providers, telecom operators, and digital service developers, while also focusing on enabling experimentation, innovation, and monetization of AI and data services.

By consolidating the essential building blocks of the 6G-DALI value network, the following canvas offers a comprehensive snapshot of the project's innovation-to-market pathway and its alignment with European digital sovereignty and data governance principles. It serves as a strategic foundation for business planning, ecosystem engagement, and long-term sustainability of the project's results.

3.3.3 Alignment with Gaia-X principles

The 6G-DALI Data Space aligns with core GAIA-X principles to enhance sustainability, scalability, and European data sovereignty. GAIA-X provides the foundational framework for federated trust and certification, positioning the 6G-DALI Data Space within the broader European data infrastructure ecosystem. This alignment ensures compliance with the Data Governance Act and supports the European Data Strategy objectives while contributing to the emerging Common European Data Spaces in telecommunications and mobility sectors.

The integration with GAIA-X principles facilitates interoperability with other sectoral data spaces and enables cross-sector innovation, particularly where 6G connectivity intersects with energy, mobility, and public services. By adopting Gaia-X guidelines, the 6G Data Space addresses the telecommunications sector's specific requirements while maintaining adherence to established European data governance standards. The alignment encompasses several key areas:

- **Data and metadata management** follows Gaia-X guidance for data identity, access control, and compliance frameworks, ensuring secure data exchange between diverse stakeholders including mobile network operators, AI developers, and vertical service providers. While Gaia-X promotes federated approaches, the DALI implementation maintains centralized data control to meet specific project requirements.
- The system utilizes standardized **data connectors** that enable secure, interoperable data exchange across different 6G systems, ensuring data discoverability and reusability for real-time applications such as predictive maintenance and network optimization.
- **Compliance and certification** mechanisms ensure that data, AI models, MLOps workflows, and DataOps processing pipelines exchanged within the 6G-DALI ecosystem adhere to Gaia-X standards for trustworthiness, security, and interoperability. This creates a standardized infrastructure that supports various 6G use cases while maintaining European data sovereignty requirements.

The Gaia-X alignment fosters collaboration among network operators, cloud providers, and AI developers while reducing traditional data silos. This framework ensures that critical data processing capabilities remain under European governance, supporting digital autonomy objectives. Additionally, the alignment creates potential pathways for exposing collected 6G data sets to the broader research and industry community, enabling wider utilization of generated research outputs within the Gaia-X ecosystem. Through this principled alignment, the 6G Data Space contributes to a trusted, interoperable data ecosystem that balances technical innovation with European values of data sovereignty and digital autonomy.

Table 4: Business Model Canvas 6G-DALI

Key Partnerships	Key Activities	Value Propositions	Customer Relationships	Customer Segments
<p>Academic and industrial research partners. Telecom operators, cloud vendors, data platform service providers and infrastructure providers. Standardization bodies (e.g., ETSI, 3GPP, CEN/CLC, ISO/IEC) and Gaia-X ecosystem. Public-private collaborations and EU-funded initiatives.</p>	<p>Development and operation of MLOps and DataOps services. Data set creation, curation, and validation. Deployment of experimentation and emulation environments. Stakeholder engagement and standardization efforts.</p>	<p>End-to-end Trustworthy AI framework combining MLOps and DataOps tailored for 6G. Access to synthetic and real data sets through a secure 6G Data Space. Digital Twin testbed for realistic and reproducible AI experimentation. Tools for service assurance, drift detection, and continuous model retraining. Federation and monetization of data via Gaia-X-aligned architecture.</p>	<p>Co-creation with verticals and AI developers. Open innovation via accessible testbeds and share-able data sets. Long-term partnerships through standardization and data governance frameworks.</p>	<p>Network/Cloud Providers (NCP): MNOs, MVNOs, Telco vendors. Infrastructure providers hosting AI models and data pipelines. Vertical Service Providers (VSP): Enterprises in domains like eHealth, public safety, and smart cities. AI Experts & ML Developers (AEMD): Technical professionals and researchers. Cloud Infrastructure Service Providers (CISP): Infrastructure operators supporting MLOps/AI. Framework Service Providers (FSP): Entities that offer framework software services. Data Providers (DP): Entities leasing infrastructure.</p>
Key Resources			Channels	
<p>Trustworthy AI-native 6G architecture with integrated Digital Twin testbed. Data Ecosystem-enabling infrastructure (6G Data Space). Expertise in telecom, AI, data engineering, ethics and regulatory framework, and network management. Consortium of leading academic and industry stakeholders.</p>			<p>6G-DALI platform access through a cloud-based portal. APIs and SDKs for developers and integrators. Workshops, technical documentation, and stakeholder engagement events.</p>	
Cost Structure			Revenue Streams	
<p>R&D and platform development. Infrastructure maintenance (e.g., testbeds, dataspace hosting). Human resources for technical and stakeholder support. Dissemination, training, and standardization activities.</p>			<p>Licensing for DataOps and MLOps platform features. Subscription-based access to data sets, tools, and testbeds. Monetization of real world and synthetic data and AI data-set-generation-as-a-service. Consultation and training for integration and deployment.</p>	

4 Description of the PoCs

This Section outlines the three Proofs of Concept (PoCs) that will be developed during the project to validate the key components of the overall solution and demonstrate its technical and business feasibility. Each one provides a description of the PoC, its participants' needs and benefits as well as its KPIs. In particular, Section 4.1 regards the first PoC “Data management and experiment on demand”, Section 4.2 is about the second PoC “AlaaS for CDN apps via cross-testbed decentralized MLOps” and finally, Section 4.3 describes the third PoC “DTT and RLOps for large and medium scale experiments”. For each PoC a set of specific KVis is defined towards the assessment of performance accurately.

4.1 PoC#1 - Data management and experiment on demand

The first PoC focuses on the data management and experiment-on-demand capabilities of the 6G-DALI framework, addressing the critical need for high-quality, accessible, and interoperable data sets to support AI and machine learning development for 6G systems, intelligent data discovery, automated data set transformation, and on-demand experimentation. It consists of two experiments, each showcasing key features of the 6G-DALI ecosystem, as described in the following subsections.

4.1.1 Description

The development of AI and machine learning models for 6G systems and services requires access to large, high-quality data sets. However, such data sets are often scarce, fragmented, or in need of cleaning and standardization to conform with widely accepted formats. In parallel, emerging initiatives like *Gaia-X* and *IDSA* are pioneering new approaches to secure and interoperable data sharing through concepts such as **Data Spaces, Catalogues, and Marketplaces**. Sectors like transportation and healthcare have already begun adopting these initiatives, demonstrating their value and practicality.

Bringing these advancements into the communications domain presents a unique opportunity to accelerate the development of robust, efficient, production-ready AI/ML frameworks for 6G. As part of this vision, the **Data Management and Experiment-on-Demand Proof of Concept (PoC)** in 6G-DALI aims to highlight the benefits of adopting these data-sharing principles. A central capability of the 6G-DALI AI framework is to respond to user requests for 6G data through an integrated and intelligent data ecosystem.

At the heart of this ecosystem is the **6G Data Space**, which interconnects the project's three 6G testbeds, each contributing unique data sets and experimentation capabilities. The PoC supports two main operational workflows:

- **Utilization of Existing Data sets:** Leveraging data sets already available in the 6G Data Space, originating from the testbeds.
- **On-Demand Data set Generation:** Executing targeted experiments on the testbeds to generate new data sets when no existing data set meets the experimenter's requirements.

A key element of this PoC is **data set discovery**—determining whether a suitable data set is already available or needs to be generated. To support this, the AI framework will leverage Data Space services developed in alignment with *Gaia-X* and *IDSA* specifications, ensuring secure, interoperable discovery and retrieval. Beyond data access, the PoC will also demonstrate how the 6G Data Space enables the creation of **ELT (Extract, Load, Transform)** pipelines. These pipelines will support the transformation of data sets according to user-defined needs, such as data cleaning or augmentation. Once processed, the enriched data sets will be reintegrated into the 6G Data Space, making them available for future reuse by other researchers. These operations are showcased in Figure 3.

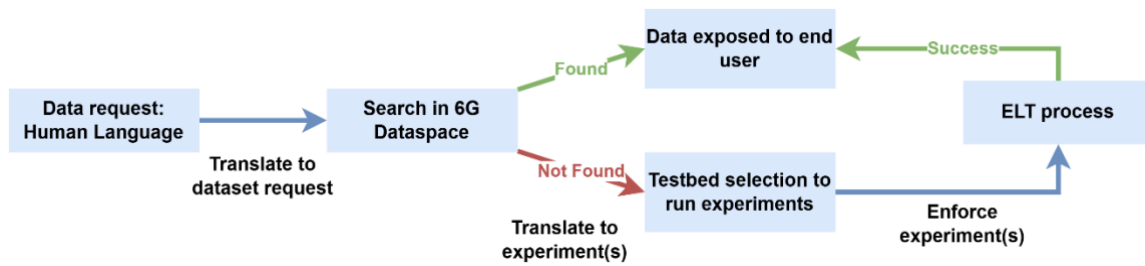


Figure 3: DataOps for existing and on demand data sets.

Through these workflows, the PoC will validate the following capabilities:

- **Integration** of the AI framework with the 6G Data Space and Gaia-X services for cataloguing, authorization, and trust management.
- **AI-driven translation** of user data requests into experiments using large language model (LLM) technologies.
- **Execution of ELT pipelines** to populate, transform, and reload data sets into the dataspace, enabling continuous data enrichment and reuse for analytics and research.

The two experiments to be performed as part of this PoC are:

- **Experiment 1.1: Data search and extraction via Gaia-X service catalogue.**

This experiment demonstrates the ability of the 6G-DALI framework to fulfil user data requests—also referred to as user intents—using data sets that are already available within the 6G Data Space, without the need to run new experiments. It showcases how the system can intelligently interpret user needs, perform secure and trusted data discovery, and apply automated data transformation procedures. This experiment validates the framework’s ability to translate abstract user intents into precise data set discovery actions, perform trusted search and access through Gaia-X compliant services, and apply transformation workflows that prepare data for immediate use in AI/ML development or analytical applications. It highlights the seamless interplay between intelligent interfaces, secure data infrastructure, and advanced data processing within the 6G-DALI ecosystem.

The experiment begins with a user submitting a data request through the 6G-DALI Northbound API. This interface integrates a large language model (LLM) that interprets the user’s high-level intent and translates it into a structured search query. The query is then executed against the 6G Dataspace’s service catalogue, which adheres to Gaia-X principles and incorporates tools such a Data and Service Catalogue, along with trust and security mechanisms for compliant and reliable data handling.

For the purposes of this demonstration, the Data Space is pre-populated with 3 curated and published data sets originating from the EUR [12] and KUL[13] [14] testbeds. Once the data set search is complete, the results are presented to the user. If a relevant data set is found, it can be used immediately or further refined. For the refinement process, the system may automatically trigger or allow the user to initiate an ETL (Extract, Transform, Load) pipelines that perform data cleaning and augmentation, leveraging algorithms developed within the 6G-DALI project to ensure the data set is fully tailored to the user's intended use.

- **Experiment 1.2: Data on-demand and enhanced ETL pipelines**

This experiment focuses on demonstrating the complete end-to-end management of a user data request that cannot be satisfied with existing data sets in the 6G Data Space. Unlike scenarios where data sets are readily available through the service catalogue, this experiment activates the second operational workflow of the PoC: the on-demand generation of data sets through targeted experimentation across the project’s testbeds. Through this experiment, the PoC demonstrates the seamless integration of intelligent orchestration, AI-powered translation of user needs into experiments, and automated ETL processes within the 6G-DALI framework. It validates how this

framework can dynamically respond to evolving data demands, promote data set enrichment, and contribute to a scalable, collaborative research environment.

The process begins when a user submits a data request. Since no matching data set is found in the Data Space, the request is automatically translated into an experiment description using a dedicated large language model (LLM), designed to interpret and convert high-level data needs into executable experiment plans. Following this translation, a testbed selection mechanism determines the most appropriate infrastructure to run the experiment based on the nature of the request. The experiment is then deployed across the available testbeds: (1) EUR for Open RAN scenarios, (2) ISI for the cloud continuum, and (3) KUL for advanced 6G wireless communication use cases. At least two of these testbeds will be engaged during executions.

Once the experiments are completed and raw data is collected, the system triggers an ETL (Extract, Transform, Load) pipeline. This pipeline is responsible for extracting the data from the testbed environments, transforming it through cleaning and augmentation methods developed in 6G-DALI, and finally loading the enriched data set back into the 6G Data Space. The result is a high-quality, reusable data set that can be accessed by other users or systems in future interactions.

Additionally, EUR & ISI will generate as part of this experiment a dataset to help train open source LLMs for chat interactions with experimentation platforms. The data set will include the full transcript of the interactions between users and the 6G-DALI chat system. This will include anonymized pairs of texts including the questions and answers used by the AI system to:

- properly search for a data set in the 6G Data Space,
- select the testbed where a data set generation experiment should be run,
- set up the experiment according to the user's needs and,
- define the DataOps pipeline to be executed on the raw collected data.

4.1.2 PoC participants

This section introduces the participants involved in the PoC, describing who they are, what benefits they expect, and what needs they have. This information helps align the PoC with stakeholder priorities and ensures that it delivers value to all parties.

4.1.2.1 PoC Participants' identification

This PoC is expected to involve the following main actors:

- **Data Platform as a Service Providers (DPaaS)**, offering infrastructure and platforms. In 6G-DALI, they support hosting of ML models, training data sets, and data-sharing environments. **(TID)**
- **AI experts and ML developers (AEMD)**, representing the stakeholders responsible for building the ML lifecycle framework to evaluate the models using the 6G-DALI experimentation capabilities. **(IQU, SPW, TUD)**
- **Data Provider (DP)**, representing the stakeholders that provide the testbed infrastructure where the experiments are deployed. **(ISI, KUL and EUR)** will provide the testbeds that will be involved in the experiments.

4.1.2.2 PoC participants' benefits

The aforementioned participants benefit from the PoC in the following ways:

- **AI experts and ML developers (AEMD)**, using the 6G-DALI's infrastructure researchers from every domain will greatly benefit from the ability to discover or create on-the-fly data sets that will help them evaluate and test their applications using real-world data in real-world conditions and using real hardware using a unified experimentation environment and automated pipelines.
- **Data Platform as a Service Providers (DPaaS)**, demonstrate value as infrastructure providers for scalable, data processing and transformation workloads across cloud-native and edge environments.

- **Data Provider (DP)**, validate their infrastructure's capability to share data sets and generate on-demand ones for the 6G domain.

4.1.2.3 PoC participants' needs

The PoC participants' needs are the following:

- **AI experts and ML developers (AEMD)**, have high needs for robust tools for AI model development, well-labelled and easily discoverable data sets, easy-to-reach execution environments and facilities.
- **Data Platform as a Service Providers (DPaaS)**, need scalable, interoperable hosting solutions for ELT workflows and demand seamless integration with data sources.
- **Data Provider (DP)**, need testing, orchestration and monitoring of their facilities.

4.1.3 KPIs Definitions

The KPIs of this PoC are the following:

- **Experiment 1.1: Data search and extraction via the Service Catalogue**

The first experiment of the PoC deals with existing data sets already loaded to the 6G Data Space. The goal is to validate the operation of the Data Space and the ability to apply DataOps efficiently to the data sets on them to generate new versions that are more compliant to the users' needs. Specifically, we will measure and validate the following KPIs:

- Reduce the time to fetch for an existing data set by half (50%) by report to the case where no 6G Data Space exists.
- 100% of data requests expressed via the intent API, based on human language.
- Showcase 100% compliance with Gaia-X.
- Apply DataOps pipelines to improve (clean or augment) 3 existing data sets. These original data sets already exist in the 6G Data Space and are the following (provided by EUR and KUL):
 - [12] from [2] the EUR testbed
 - [13] and [14] from the KUL testbed
- **Experiment 1.2: Data on-demand and enhanced ELT pipelines**

The second experiment of the PoC deals with the need to generate new data sets from the testbeds connected to the 6G Data Space. The goal of this experiment is to validate the ability to generate new data sets through the execution of experiments in the EUR, ISI and KUL testbeds through the designed testbed adapters, as well as the registration of new data sets to the 6G Data Space. Finally, the PoC will also demonstrate how DataOps pipelines can be applied to newly reported data sets being added to the Data Space. Specifically, we will measure and validate the following KPIs:

- Improve the quality of collected data set using the data cleaning algorithms by 30% compared to the case where no algorithms are used.
- Generate 3 new open data sets one from each of the connected testbeds and apply data cleaning and improvement on the generated data sets:
 - O-RAN data set from the EUR testbed,
 - Cloud Continuum data set from the ISI testbed,
 - 6G Wireless Communications data set from the KUL testbed
- Translate 100% of data requests to experiments, for data generation purposes (i.e., no pre-existing data sets)
- Satisfy at least 70% of data provisioning requests via the natural language-based APIs.
- Publish an open dataset with interactions of users with the 6G-DALI LLM powered chat system.

4.1.4 KVis definitions

This section provides a definition of the KVis for each experiment under PoC#1.

- **Experiment 1.1: Data search and extraction via the Service Catalogue**

- User satisfaction from intent-based data requests upper to 90%. Initial user requests will be validated through surveys at the end of each experiment.

User satisfaction of an intent can be defined as the degree to which a system's response aligns with the user's expectations for that intent. In this context, a user intent represents the goal the user wishes to achieve through a request, while the system's role is to correctly interpret that intent and translate it into concrete actions. Examples of possible intents include, for instance, retrieving mobility data from the EUR testbed for analysis or searching for energy consumption datasets from trusted sources.

The level of satisfaction depends on how well the performed actions match the original intent. This can be assessed through multiple dimensions:

- Accuracy – did the system correctly understand the intent?
- Relevance – was the response aligned with the user's actual need?
- Completeness – did the system provide all the necessary information or fully complete the requested task?
- Ease of use – was the interaction smooth, efficient, and free of unnecessary effort?
- Outcome success – does the user perceive that their goal has been achieved?

As satisfaction is inherently subjective, this KVI will be measured through post-experiment surveys, where users evaluate how effectively the system addressed their requests across the dimensions above. Achieving a satisfaction rate above 90% will indicate that the 6G-DALI framework is capable of delivering trustworthy, intent-driven data discovery and preparation.

- 100% compliance with relevant ethical and legal requirements. Based on that:

Compliance with all relevant legal requirements as per EU AI Act shall reach 100%, as evidenced in terms of evaluation metrics, compliance checklists, and documented evaluations that operationalise the criteria of traceability, adequacy, and accountability through assessment within the experiment of validation of the Data Space and applying DataOps to generate improved data set versions that better meet user needs.

Traceability refers to the ability to reconstruct and verify every step in the fulfilment of a user intent, from request interpretation through dataset discovery and transformation to delivery. This ensures that provenance, decision points, and applied safeguards can be followed in a transparent and auditable way.

Adequacy addresses the alignment between system outcomes, user intent, and applicable ethical and legal requirements. It ensures that responses are not only technically correct but also contextually appropriate, explainable, and compliant with frameworks such as GDPR and the AI Act.

Accountability guarantees that responsibilities, roles, and decision-making processes are clearly documented, with records that can be verified internally or externally, thereby enabling meaningful oversight and governance.

Evidence for these criteria will be provided through several complementary instruments. Evaluation metrics capture quantifiable aspects of compliance, for example the percentage of data requests that result in GDPR-compliant outputs. Compliance checklists provide structured validations, adapted from ETHAI's operationalisation of ethical principles, to confirm whether each requirement has been addressed during system operation. Documented evaluations take the form of formal reviews and reports, demonstrating that processes and outcomes are systematically assessed against GDPR, the AI Act, and project-specific governance standards.

- **Experiment 1.2: Data on-demand and enhanced ELT pipelines**

- User satisfaction upper to 90%. Initial user requests will be validated through surveys at the end of each experiment.

Here user satisfaction is defined similarly than for experiment 1.1 but with a relevant difference: unlike experiment 1.1, where satisfaction depends primarily on the system's ability to find, retrieve, and prepare existing datasets from the 6G Data Space, in experiment 1.2, user satisfaction is defined as the degree to which the system successfully delivers a new, high-quality dataset that directly responds to the user's unmet request. Thus satisfaction in this case hinges on the on-demand generation process: whether the LLM correctly translates the user's request into an experiment description, whether the appropriate testbeds are selected and orchestrated, and whether the resulting dataset is enriched and reusable.

Accordingly, satisfaction will not only reflect accuracy, relevance, and ease of interaction, but also the timeliness, completeness, and quality of the newly produced dataset, together with the user's perception that their evolving data needs have been met through a scalable, collaborative workflow.

- 100% compliance with relevant ethical and legal requirements

In Experiment 1.1, compliance measurement focuses on the use of existing datasets: whether their discovery, access, and transformation are handled in line with GDPR, the AI Act, and project-specific governance requirements. The emphasis is on provenance, lawful access, data quality during transformation, and the transparency of how user intents are matched with available datasets.

In Experiment 1.2, by contrast, compliance measurement must also extend to the generation of new datasets through targeted experiments across the testbeds. This introduces additional points of scrutiny: whether the experimental setup itself respects ethical and legal requirements; whether new data is collected, processed, and enriched in a way that safeguards privacy, fairness, and data protection principles; and whether documentation of the newly created datasets provides sufficient traceability and accountability for future reuse. A key additional concern is the risk of bias: newly generated datasets may reproduce or amplify existing imbalances. For this reason, the project will consider suitability assessments for new datasets, with attention to risks such as bias or representational imbalance, and outcomes will be documented where appropriate to support transparency and accountability.

Compliance in this experiment will therefore be evidenced not only through the use of existing governance checks, but also through mechanisms adapted to dataset generation. These may include compliance verification against GDPR, the AI Act, and project-level governance requirements; suitability assessments addressing issues such as bias; and processes to document dataset provenance, including collection methods, applied transformations, and responsible actors.

The difference, therefore, lies in scope: while Experiment 1.1 validates compliance in managing what already exists, Experiment 1.2 requires validating compliance in bringing new data into existence, with particular attention to bias risks and the ethical-legal soundness of the generation process from the outset.

- Alignment of experiment results with users' objectives and priorities, such as timeliness of delivery, relevance of the generated dataset, completeness of the information provided, and perceived quality of enrichment and transformation processes, shall exceed 90%, as validated through user surveys or other social sciences techniques conducted after the experiment of generation of new data sets from the testbeds connected to the 6G Data Space.

In Experiment 1.2, user satisfaction and alignment with objectives cannot be measured solely by whether a dataset is returned, but rather by whether the newly generated dataset meaningfully meets user priorities. These priorities include the relevance of the dataset to the

original request, the completeness and usability of the collected data, the quality of enrichment through ETL pipelines, and the timeliness of delivery. Unlike Experiment 1.1, where alignment is mostly about the discovery of suitable existing datasets, here satisfaction reflects the system's ability to translate user needs into experiments, orchestrate testbeds effectively, and return enriched datasets that are ready for use in research or AI/ML development. Surveys and other social sciences techniques will validate whether these outcomes correspond to user expectations, with a target of exceeding 90% alignment.

4.2 PoC#2 - AIaaS for CDN apps via cross-testbed decentralized MLOps

The second PoC focuses on the decentralized MLOps capabilities of the 6G-DALI framework, because of the increasing need for robust, automated solutions that can manage machine learning workflows across diverse infrastructures. It consists of two experiments described in detail in the following subsections.

4.2.1 Description

The 6G-DALI cross-testbed decentralized MLOps PoC is designed to validate the end-to-end AI framework with a focus on decentralized MLOps, meta-orchestration, and AI experimentation services. The PoC targets the evaluation of cooperative and distributed mechanisms for managing machine learning (ML) models and aims to demonstrate their effectiveness within a vertical Content Delivery Network (CDN) use case. In particular, it assesses the framework's ability to automate the entire ML lifecycle, covering training, deployment, inference, hyperparameter optimization and transfer learning, without manual intervention. A key goal is to demonstrate that 6G-DALI can operate seamlessly across different testbeds and MLOps stacks, while being able to support distributed AI/ML tasks and workloads.

The motivation behind this PoC is that despite the growing interest in MLOps for automating ML workflows, current solutions often require manual steps at various stages, especially when deployed in complex or production environments. These limitations hinder full automation and scalability, which are essential for realizing the potential of AI in the dynamic, heterogeneous infrastructure characteristics of beyond 5G and emerging 6G networks. Existing MLOps frameworks tend to be fragmented to specific use cases, lacking the flexibility to support diverse ML techniques and deployment scenarios across the cloud-to-edge continuum. The 6G-DALI PoC addresses these gaps by aiming to offer a unified, adaptable solution capable of overcoming interoperability challenges and enabling consistent AI operations across distributed systems. The PoC consists of two experiments:

- **Experiment 2.1: Hyperparameter optimization and placement of ML models at the Cloud-Edge-Continuum** The first experiment will showcase how the experimenters and users can leverage the 6G-DALI framework for efficient ML model training, optimization and deployment across the Cloud-Edge-Continuum. More specifically, the users can run complex ML workflows, including training, fine-tuning, transfer learning, and model distillation, using the integrated Hyperparameter Optimization (HPO) service, while the system intelligently allocates the minimal computing resources required to meet performance goals like low latency or cost-efficiency. This is particularly valuable for deploying AI workloads on constrained edge or deep-edge devices, where hardware limitations are a key challenge. Users benefit from streamlined experimentation that automatically adapts to infrastructure constraints and optimizes resource use without the need for manual tuning. Additionally, the experiment showcases the use of model compression and quantization, allowing users to deploy powerful AI models even on limited-capacity devices. Overall, the experiment aims to double the number of supported ML workloads using AutoML-driven HPO, ensure deployment of at least one ML workload at a constrained edge device, and generate one data set as part of the evaluation.
- **Experiment 2.2: ML model Benchmarking and drift detection during vertical application testing** In the second experiment, the PoC empowers experimenters and verticals to validate, test, and benchmark AI/ML models in staging environments before full production deployment. This

capability is demonstrated through a real-world application using ICOM's commercial future-proof CDN (fs|CDN) platform. More specifically, using the 6G-DALI MLOps framework, users can evaluate the performance of AI models that predict resource demand and optimize content delivery, enhancing the Quality of Experience (QoE) for end-users. Experimenters can deploy these models across various cloud-native and edge infrastructures (regardless of ownership or capacity) and use pre-trained models from the ML Catalogue to dynamically control video playback and transcoding settings. Furthermore, the experiment integrates trustworthy AI capabilities, enabling observability and continuous monitoring of AI model behaviour. Users benefit from automated drift detection, which can trigger model retraining or adaptation to maintain accuracy and reliability over time. This ensures that services stay responsive and resilient in dynamic network conditions. The inclusion of data augmentation techniques also helps improve model robustness, allowing users to prepare AI applications for diverse and unforeseen scenarios. Overall, this scenario provides a practical pathway for bringing AI-enhanced services into production, supported by continuous performance monitoring and adaptability features.

4.2.2 PoC participants

This Section outlines the participants in the PoC, detailing their identities, expected benefits, and specific needs. Understanding this helps ensure the PoC is aligned with stakeholder interests and delivers meaningful value to everyone involved.

4.2.2.1 PoC Participants' identification

This PoC is expected to involve the following main actors:

- **Vertical Service Provider (VSP)**, representing the stakeholder that delivers domain-specific cases. In the second experiment of this PoC, CDN platform is provided. **(ICOM)**
- **AI experts and ML developers (AEMD)**, representing the stakeholders that deliver the MLOps solution and create and optimize the machine learning models through hyperparameter optimization and transfer learning. **(IBM, NXW and KUL)**
- **Data Provider (DP)**, representing the stakeholders that provide the testbed infrastructure where the experiments are deployed. **(KUL and EUR)**
- **Data Platform as a Service Provider (DPaaS)**, representing the stakeholder that host the 6G-DALI framework including ML models catalogue and transformation services and the stakeholder that offers cloud infrastructure resources. **(TID, ATH/ISI)**

4.2.2.2 PoC participants' benefits

The aforementioned participants benefit from the PoC in the following ways:

- **Vertical Service Providers (VSP)** will gain predictive capabilities for content/resource management and will benefit from MLOps-driven automation and QoS optimization in their services.
- **AI experts and ML developers (AEMD)** will leverage the AutoML-driven HPO services to successfully deploy models at constrained edge devices.
- **Data Providers (DP)** validate their infrastructure's capability with heterogeneous MLOps stacks.
- **Data Platform as a Service Provider (DPaaS)** demonstrate value as infrastructure providers for scalable, distributed AI workloads across cloud-native and edge environments.

4.2.2.3 PoC participants' needs

The PoC participants' needs are the following:

- **Vertical Service Providers (VSP)** need to use 6G-DALI as a staging/testing framework that mimic production CDN use case in order to compare various models' performances and ensure continuous monitoring of the deployed model's performance.
- **AI experts and ML developers (AEMD)** need advanced tools for model lifecycle management (training, tuning, inference) to support their activities.

- **Data Providers (DP)** need testing, orchestration and monitoring of their facilities.
- **Data Platform as a Service Provider (DPaaS)** need scalable, interoperable hosting solutions for diverse AI workflows and demand seamless integration with diverse MLOps stacks.

4.2.3 KPIs Definitions

The KPIs of this PoC are the following:

- **Experiment 2.1: Hyperparameter optimization and placement of ML models at the Cloud-Edge-Continuum**
 - Double the number of ML workloads supported via HPO driven by AutoML.
 - Deploy at least 1 ML workload at a constrained edge device.
 - Generate 1 data set: Resource prediction, with timeseries data leveraged for microservice placement at the Cloud-Edge-Continuum will be collected from the Cognitive Continuum enabler at the ISI testbed, with CPU, RAM, pod placement data collected via k8s telemetry from all k8s clusters and microservices deployed at the Cloud-Edge-Continuum. Transport Network telemetry (e.g., TWAMP, IP Path Management, Network Load) that may affect pod placement will also be included.
- **Experiment 2.2: ML model Benchmarking and drift detection during vertical application testing**
 - Have at least 0.5 MOS score gain for the CDN via resource prediction.
 - Reduce at least by 20% vertical SLA breaches via trustworthy AI techniques such as drift detection and conformal prediction.
 - Improve by at least 20% prediction accuracy for the PHY model in unseen scenarios, via Trustworthy AI approaches that improve model generalization ability via data augmentation.
 - Generate 2 data sets: i) **KUL** will extend its existing ultra-dense PHY dataset with more than 200000 datapoints [13] to offer technical trustworthiness of AI models for PHY trained with this dataset. The dataset will be augmented with experiments in the same environment with (synthetically) blocked antenna elements to allow trained AI models to generalize in many PHY layer scenarios. Various other augmentation methods will be considered, such as hardware impairments, or clock offsets. ii) **ICOM** will provide a labeled data trace with the CDN performance logs and the respective configurations, deployment and pod placement parameters.

4.2.4 KVIs Definitions

This section provides a definition of the KVIs for each experiment under PoC#2.

- **Experiment 2.1: Hyperparameter optimization and placement of ML models at the Cloud-Edge-Continuum**
 - Achieve at least 90% reduction in energy consumption through model compression and quantization, with no more than 5% accuracy loss compared to the FP32 precision model, demonstrating efficient ML model training, optimization, and deployment across the Cloud-Edge-Continuum.

This KVI highlights the societal and ethical relevance of energy-efficient AI in 6G contexts. While Key Performance Indicators (KPIs) focus on technical achievements such as latency or accuracy, this KVI captures the value dimension: demonstrating that advanced ML workflows can be made more sustainable. By targeting a 90% reduction in energy consumption through compression and quantization - with only a marginal ($\leq 5\%$) loss in accuracy compared to full-precision models - the experiment directly addresses concerns over the environmental footprint of AI.

The indicator matters because training and deploying ML models across the Cloud-Edge-Continuum has escalating energy demands, contributing to CO₂ emissions and straining

hardware resources. Validating that significant energy savings can be achieved without compromising usability provides evidence that AI in 6G can align with EU priorities on the Green Deal, digital sustainability, and resource efficiency. Measurement will be based on systematic benchmarking of compressed versus baseline models, capturing both energy consumption and accuracy, but the true value lies in showing that environmental responsibility and technological performance can be balanced.

- Demonstrate accuracy, reliability, robustness, accountability, transparency of ML workload operations through the ETHAI assessment report.

To clarify the scope of this KVI, the following section explains in detail each of the concepts it encompasses. This makes explicit how accuracy, reliability, robustness, accountability, and transparency are understood within the ETHAI framework and how they will be assessed in practice.

Accuracy: Within ETHAI, accuracy refers not only to the technical precision of ML workload outputs but also to their adequacy in addressing the user's intended purpose. Accuracy is assessed in context, ensuring that performance metrics are meaningful for the specific use case rather than treated as abstract figures.

Reliability: Reliability refers to the degree to which the outputs of ML workloads can be trusted as dependable and reproducible. In ETHAI, this involves validating that results are consistent over time, across repeated executions, and against reference benchmarks, giving users confidence that the system behaves as expected and its outcomes can be relied upon for decision-making. ETHAI emphasizes iterative testing and validation cycles, ensuring that reliability is demonstrated not just in controlled settings but in realistic, heterogeneous environments.

Robustness: Robustness reflects the resilience of ML workloads against perturbations, failures, or adversarial conditions. ETHAI operationalizes this concept by linking it to explicit requirements derived from stakeholder input and by embedding safeguards into system design, ensuring robustness against domain-specific risks and that it is not accidental but systematically achieved.

Accountability: Accountability in ETHAI is grounded in traceability and responsibility allocation. Every decision, transformation, or model adjustment within ML workloads should be documented, with clear attribution of roles and justifications, enabling oversight, contestability, and compliance with regulatory frameworks such as the AI Act.

Transparency: Transparency is treated by ETHAI as both a technical and ethical principle: it requires that the processes behind ML workloads - from data handling to decision-making - are intelligible to stakeholders. This is achieved through explanation adequacy criteria tailored to different audiences ensuring that transparency is not generic but contextual and actionable.

The assessment of these concepts will be carried out through a combination of compliance checklists, structured assessment criteria, formal reviews, and dedicated workshops with stakeholders and project partners. Together, these methods ensure that the ETHAI assessment report provides a systematic and verifiable account of how accuracy, reliability, robustness, accountability, and transparency are demonstrated in practice.

- **Experiment 2.2: ML model Benchmarking and drift detection during vertical application testing**
 - Demonstrate accuracy, reliability, robustness, accountability, transparency of AI/ML models through the ETHAI assessment report.

Although the core concepts of accuracy, reliability, robustness, accountability, and transparency are defined consistently across experiments, their operationalization in

Experiment 2.2 differs from Experiment 2.1 because of the context of benchmarking, drift detection, and lifecycle assurance in a vertical CDN application. In this setting, the emphasis shifts from efficient deployment to ensuring trust and resilience of models over time in dynamic, production-like conditions. Following, the definition of key concepts adapted for experiment 2.2:

Accuracy: In this experiment, accuracy is assessed not only in terms of baseline model performance but also in terms of its preservation over time. Drift detection mechanisms ensure that accuracy does not silently degrade as data distributions and network conditions evolve.

Reliability: Reliability here refers to the degree to which model predictions and optimisations for content delivery can be consistently trusted in production-like environments. It highlights the dependability of results in sustaining service quality across variable infrastructure and workload conditions.

Robustness: Robustness in Experiment 2.2 includes the system's capacity to handle distribution shifts, novel scenarios, and adversarial conditions without unacceptable performance loss. Data augmentation strategies contribute to strengthening this resilience.

Accountability: Accountability is demonstrated through systematic logging of model behaviour, drift events, and retraining actions. This documentation ensures that the lifecycle of models can be audited, making it possible to verify whether appropriate corrective measures were taken when accuracy or performance declined.

Transparency: Transparency here takes the form of observability: developers and operators must be able to monitor, interpret, and validate drift detection outcomes and retraining decisions. This ensures that adaptive behavior in the CDN use case can be scrutinized and trusted.

- Compliance with all relevant legal requirements shall reach 100%, as evidenced through verification and documentation within the experiment of benchmarking of ML models in staging environments before full production deployment.

This KVI ensures that all ML model operations in Experiment 2.2 comply fully with relevant legal requirements, with particular attention to privacy and data protection duties under the GDPR and the EU AI Act. In Experiments 1.1 and 1.2, compliance focused on datasets — either retrieving or generating them, and ensuring obligations in data handling. By contrast, in Experiment 2.2 the emphasis shifts to ML model operations in a vertical CDN application, where privacy is especially critical. User data may be processed for training, benchmarking, or drift detection, and outputs directly affect end-user experience.

Compliance will therefore be assessed not only at the point of project setup but also throughout the entire lifecycle of model operations, ensuring that user data is processed lawfully, securely, and with adequate safeguards. Evidence will be gathered through compliance checklists, structured reviews, and documentation of privacy-preserving practices, providing assurance that model benchmarking and monitoring respect both technical and legal standards from the outset.

- Demonstrate absence of bias in terms of data distribution, representational balance, and performance across heterogeneous contexts in the specific setting where the AI/ML models are applied through the ETHAI assessment report.

To make the KVI precise and meaningful, bias is interpreted here in relation to the dimensions most relevant to AI/ML in the CDN and 6G context. Since ETHAI emphasizes context-sensitivity, this is not treated as a one-size-fits-all problem but rather as one that must be justified by the use case. For Experiment 2.2, meaningful bias dimensions include: data distribution and representational imbalance (e.g. training data not reflecting diverse network or traffic conditions), performance disparities across contexts (e.g. model accuracy varying

significantly by device, region, or workload type), and systematic favouring of certain users or content types (e.g. unfair allocation of resources in CDN optimization). ETHAI will operationalize these dimensions through assessment tools such as structured checklists, comparative benchmarks, and stakeholder validation, ensuring that the absence of bias is documented and verified in the assessment report.

A further analysis on KVI's will be conducted in deliverable D2.4

4.3 PoC#3 - DTT and RLOps for large and medium scale experiments

This section presents the PoC DTT and RLOps for medium- and large-scale deployments, showcasing 6G-DALI's ability to conduct extensive experiments through its AI framework and digital twin infrastructure.

4.3.1 Description

The growing need to develop machine learning models tailored for 6G networks is significantly constrained by the lack of representative data sets, particularly when it comes to modelling the RAN and CN. This limitation arises not only from the scarcity of publicly available data sets but also from the inadequacy of the scenarios used to generate them. Most existing data sets fail to capture the scale and complexity expected in 6G environments, such as a high density of UE per cell, large-scale multi-cell deployments, and realistic mobility and handover scenarios. While many existing platforms like those introduced by SNS-C projects (6G-BRICKS and SUNRISE-6G) facilitate experimentation with different configurations for 6G RAN and CN, they are typically limited in the number of UEs that can be simulated or emulated. This restricts the ability to evaluate critical features such as Xn handovers (mobility between gNBs via the Xn interface without involving the core network) and F1 handovers (mobility between distributed units under the same gNB central unit using the F1 interface), as well as to assess the scalability of CN functions and algorithms designed to manage dense UE populations. Furthermore, many of these platforms do not support the execution of machine learning models or reinforcement learning agents directly on the testbed. This hinders the validation of such models in realistic, scalable environments—an essential requirement for ensuring their practical applicability in 6G networks. Moreover, many platforms do not support the execution of ML models and RL agents, which require validation in realistic and scalable environments that reflect genuine deployment scenarios.

Knowing these limitations, 6G-DALI devises a novel Digital Twin Testbed (DTT) for the 6G RAN and CN. The DTT will allow the running of experiments and, hence, dat a set generation for scenarios that require a high-scale deployment running a high number of mobile UEs, several gNB cells, different RAN configurations, etc.

This DTT testbed is essential for evaluating next-generation network technologies before they are deployed in the real world. The DTT enables large-scale experimentation by simulating thousands of mobile devices and network elements such as base stations and core networks. This allows researchers and engineers to test how different network configurations perform under various conditions, including high traffic loads and dynamic user movements. The insights gained from these experiments help shape more efficient, reliable, and adaptable 6G networks. The testbed utilizes both open-source and commercial tools. It leverages the well-established OAI 5G tools in EUR and testing tools from VIA that validate how well different network components work together. These tools can simulate real-world usage, check for compliance with industry standards, and ensure that the network performs well under pressure. Additionally, the system is designed to monitor and adjust network operations intelligently. It uses specialized applications that gather performance data and make real-time decisions to optimize how the network functions. This dynamic control is a key element for future smart networks that can adapt to changing conditions instantly.

The DTT setup includes three main components:

- A configuration manager to define the experiment, such as network layout, device behaviour, and traffic patterns.

- An intelligent control layer that monitors network performance and makes adjustments as needed.
- An adapter that connects all components to a broader AI framework, enabling automated experimentation and data collection.

All experimental data is stored and shared via a centralized 6G data platform, allowing for further analysis and training of AI models. This collaborative and automated approach accelerates the development of innovative 6G technologies and ensures they are tested rigorously before real-world implementation.

For PoC#3 of 6G-DALI which will be executed on DTT, we envision two scenarios: (i) large-scale experimentation for RLOps; (ii) medium-scale DT experimentation for O_RAN. Both scenarios have common objectives, which demonstrate:

- The automatic deployment of large-scale experimentation on the 6G-DALI from the definition, deployment, and generation of data sets involving the AI framework components: user gateway, data acquisition, data cleaning, adaptation layer, and DTT adapter.
- The configuration and data set collection using the DTT and xApp, as well as the application run on top of Network Data Analytics Function (NWDAF).
- The DT capabilities to generate medium and large-scale experimentations with low and high-fidelity DT.
- The integration of the DTT via ELT to the 6G Data Space.
- RLOps mechanisms devised in 6G-DALI.

- **Experiment 3.1: Large scale experimentation for RLOps**, noted also as large-scale and low-fidelity, aims to experiment with scenarios that involve deploying a high number of UEs connected to different cells while supporting different mobility patterns. In this scenario, the UEs and RAN are based on VIA tools, while the RAN Intelligent Controller (RIC) and CN are based on OAI. This scenario has a lower fidelity as no radio is used. However, this scenario support large scale experiments, by enabling the collection of data on user mobility, high number of attach requests, and high traffic patterns, which can be used to train RL agents that optimize the system performances, such as mobility and handover procedures, dimensioning of CN functions, etc. The RL agents are expected to be on-boarded as xApp on top of RIC or an application running on top of the NWDAF. This will pave the way to showcase 6G-DALI RLOps system in terms of its Simulation-to-Reality (SIM2REAL) generalization capabilities.

This scenario has these specific objectives:

- Demonstrate the DTT capabilities to generate large-scale experiments.
- Demonstrate RLOps system of 6G-DALI.
- Showcase 6G-DALI ELT capabilities.
- **Experiment 3.2: Medium-scale DT experimentation for O-RAN.** The second scenario focuses on testing with a moderate number of real user devices using actual radio signals. This setup uses equipment from VIAVI to simulate realistic mobile network conditions. Unlike the first scenario, which was fully virtual, this one includes real radio hardware to closely mirror real-world deployments. The goal is to collect valuable data on how the network performs across different parts, including user connections, central network behaviour, and how devices move between network areas. This type of data is not yet available in O-RAN environments, as the technology is still new and current testing platforms are limited. To support this, custom applications will automatically collect and organize the data during the experiment as part of the 6G-DALI project. This will help demonstrate what is possible with more advanced and realistic O-RAN testing.

The specific objectives of this scenario are:

- Demonstrate the DTT capabilities to generate medium scale experiment with high fidelity targeting complex scenario such as O-RAN.
- Showcase ELT capabilities.

- Demonstrate experiment on demand using DTT.

4.3.2 PoC participants

This section identifies the PoC participants, along with their expected benefits and specific needs.

4.3.2.1 PoC Participants' identification

The successful execution of PoC#3 within the 6G-DALI framework relies on the collaboration of several key participant roles. Each plays a distinct function to support the development, deployment, and evaluation of both large-scale RLOps experimentation and medium-scale DT experimentation for O-RAN. The roles are as follows:

- **AI Experts and ML Developers (AEMD):** Design, implement, and tune RL agents. They integrate AI framework components to automate experimentation and ensure seamless data processing. Delivery of the HPO solution representing expertise in developing AI tools, and the MLOps solution. Contribution to AI model training, particularly in the area of physical layer (PHY) models and data augmentation techniques.
- **Data Providers (DP):** Provide and manage the physical and virtual infrastructure for experimentation. This includes setting up the OEM DTT, enabling xApp deployment, and supporting the execution of scenarios via NWDAF for data collection and real-time feedback.

4.3.2.2 PoC participants' benefits

AEMDs benefit from an automated experimentation pipeline, faster model validation cycles, and access to rich, high-quality data sets for training and evaluation.

DPs strengthen the value of their testbed infrastructure by supporting cutting-edge use cases, increasing visibility, and fostering collaboration across the 6G ecosystem.

4.3.2.3 PoC participants' needs

AEMDs require robust access to AI tooling, clean and well-labelled data sets, and reproducible experimentation environments that facilitate continuous development.

DPs need flexible infrastructure capable of supporting low- and high-fidelity Digital Twin deployments, along with tools for real-time monitoring and orchestration.

4.3.3 KPIs Definitions

The KPIs of this PoC are the following:

- **Experiment 3.1: Large-scale experimentation for RLOps**
 - Reduce by 50% the time required to validate a reinforcement learning agent compared to scenarios without Digital Twin Technology.
 - Ensure that 80% of experiments are executed automatically (no human in the loop).
 - Generate at least two large data sets tailored for large-scale scenarios:
- **Experiment 3.2: Medium-scale Digital Twin experimentation for O-RAN**
 - Ensure that 80% of experiments are automated (no human in the loop).
 - Generate at least two large data sets for medium-scale scenarios, fully compliant with O-RAN specifications.
 - Ensure that 100% of the generated data is shared with various stakeholders via the 6G Data Space.

Datasets for Experimental Scenarios 1 and 2: They will cover large and medium scalable scenarios on the RAN, to be published in open-access format like previous EUR initiatives [15] . These datasets will include:

- Data from the O-RAN fronthaul, backhaul as well as F1 interface, in addition to data regarding F1 mobility as well as Xn mobility between CU.
- Data from the Core Network: Users attached, PDU, Slice, DNNs, NSSAIs, attached subscribers, bearer context, N3 counters per attached subscriber, DNN AMBR, UE Connected
- Fault management Data: RAN, CN alarms generated

As large-scale scenarios we consider up to 100 UE, as medium scale scenarios we consider up to 10 UEs.

4.3.4 KVIs definitions

This section provides a definition of the KVIs for each experiment under PoC#3.

- **Experiment 3.1: Large-scale experimentation for RLOps**
 - Compliance with all relevant legal requirements shall reach 100%, as evidenced through verification and documentation in the context of large scale experimentation for RLOps

In Experiment 3.1, compliance with legal requirements — particularly on privacy — takes a distinct form compared to other experiments. Since the scenario is based on large-scale, low-fidelity simulations without the use of real radio signals or direct user data, the immediate risk to personal data protection is minimal. Nevertheless, compliance requires that any datasets or models used to generate synthetic mobility or traffic patterns are lawfully sourced, anonymised where relevant, and transparently documented. The KVI thus reflects the commitment to 100% legal compliance even in synthetic environments, ensuring that simulated data does not inadvertently compromise privacy and that the assumptions underlying large-scale experiments remain accountable and auditable.
 - Demonstration of accuracy, reliability, robustness, accountability, and transparency for large-scale experimentation for RLOps, by completing an ETHAI assessment report.

In Experiment 3.1, the demonstration of accuracy, reliability, robustness, accountability, and transparency must be interpreted in the context of large-scale, low-fidelity simulations. The ETHAI assessment report will not only evaluate technical performance but also document how simulation results can be trusted, reproduced, and responsibly applied to real-world scenarios. This includes verifying accuracy with respect to SIM2REAL transferability, checking reliability through repeated runs, assessing robustness under varied traffic and mobility conditions, and ensuring that assumptions underlying synthetic data generation are traceable and transparent. By embedding accountability into the experimental design, the ETHAI assessment guarantees that large-scale synthetic testing contributes responsibly to the development of RLOps without overstating its applicability.
- **Experiment 3.2: Medium-scale DT experimentation for O-RAN**
 - Compliance with all relevant legal requirements shall reach 100%, as evidenced through verification and documentation in the context of medium-scale DT experimentation for O-RAN.

In Experiment 3.2, compliance with relevant legal requirements - particularly GDPR and the AI Act - must be demonstrated in the context of medium-scale DT experimentation with O-RAN, where real radio signals and device interactions are involved. Unlike Experiment 3.1, which relies entirely on synthetic data, this scenario introduces privacy and security considerations linked to the collection and processing of realistic mobility and network performance data. The KVI reflects this by requiring 100% compliance, evidenced through systematic verification and documentation of how experimental data is sourced, anonymised where necessary, and safeguarded throughout its lifecycle. Compliance checks will ensure that sensitive information is neither exposed nor misused, and that data handling aligns with both legal obligations and responsible research standards.

- Demonstration of accuracy, reliability, robustness, accountability, and transparency for medium-scale experimentation for O-RAN, by completing an ETHAI assessment report.

In Experiment 3.2, the demonstration of accuracy, reliability, robustness, accountability, and transparency is carried out in the context of medium-scale O-RAN experimentation with real radios and devices. The ETHAI assessment report will evaluate accuracy in terms of how well models and experiments capture real-world network performance; reliability in producing consistent outcomes across repeated tests under varying radio conditions; and robustness in maintaining performance despite environmental variability, device mobility, or signal fluctuations. Accountability requires clear documentation of the experimental setup, data collection methods, and decision points, ensuring that the results can be reproduced and audited. Transparency is achieved by making the processes of data gathering, orchestration, and evaluation intelligible to developers, enabling them to scrutinise and trust the experiment’s findings.

Through this assessment, Experiment 3.2 demonstrates not only technical soundness but also the responsible use of realistic O-RAN test environments, ensuring that insights derived from medium-scale experimentation are dependable, reproducible, and aligned with ethical and legal standards.

A further analysis on KVIs will be conducted in deliverable D2.4 where the ETHAI model, the assessment methodology and the related requirements will be presented.

4.4 Sustainability Analysis

In addition to the above defined KPIs and KVIs, 6G-DALI, based on the 3 PoCs outcomes, will also perform a **sustainability analysis** of the e2e AI framework. Sustainability will be measured under 3 angles (stemming from HEXA-X-II²⁴), covering societal, economic and environmental aspects. The quantitative evaluation methodology for sustainability is outlined in Table 5.

Table 5: 6G-DALI framework quantitative evaluation of Sustainability

Relevance	Aim	How to measure it
6G networks should be sustainable from a wider socio-economic and environmental perspective.	Minimize resource consumption, i.e., the use of basic resources per bit should be minimized (notably due to Native AI integration)	<p>Societal: by reducing by >50% the (computational) resource utilization for training AI models thanks to the 6G-DALI shared pool of pre-trained models. Baseline: Resource use for training the models from scratch</p> <p>Economical: by reducing the OPEX of data generation by >20% due to intent-driven DataOps. Baseline: OPEX for data generation based on manual processes.</p> <p>Environmental by achieving >90% energy savings in inference operations via offering model quantization as a service, and allowing experimenters to represent</p>

²⁴ <https://hexa-x-ii.eu/>

		model weights with lower precision to sufficiently compress their models in a way that they fit constrained and low-power devices. Baseline: representing models with FP32 precision, when this service is unavailable.
--	--	---

4.5 Requirements Elicitation Process

The 6G-DALI requirements elicitation process was based on a discussion led by each PoC leader within each specific PoC. According to these moderate discussions, feedback was received and consolidated at Table 6 forming an overall list of General User Requirements. The same general list is leveraged at the following stages, where a fictitious user i.e. a persona profile will be created and subsequently selects his/her requirements from the general bucket list of requirements.

Table 6: General use cases requirements

Requirement	Description
REQ 1 Model Benchmarking	Enables benchmarking of pre-trained models available in the ML catalogue.
REQ 2 Model Compression	Enables model compression to fit them within resource constrained devices. Includes configurable parameters such as the compression technique (pruning / quantization / knowledge distillation / a combination from the catalogue), the compression level (manually defined / based on target accuracy / auto-optimized), the validation strategy (available in the catalogue / custom-defined / automatically selected based on the model and task), and the evaluation metric (available in the catalogue / custom-defined / automatically-selected).
REQ 3 Model Drift Detection / Continuous Monitoring	Enables continuous monitoring of performance of deployed ML models to detect any degradation. Includes configurable parameters for the drift detection process such as the drift metric (statistical distance measure / distribution shift metric / learned or heuristic-based metric) and the detection frequency (real-time / batch / daily). In case of deviation, either from a performance threshold or from a baseline model, model re-training is triggered to enhance robustness.
REQ 4 Data Space Integration	Enables seamless interoperability with the 6G Data Space, supporting data set discovery, retrieval, publication, and metadata exchange using standardized interfaces.
REQ 5 Natural Language Query Translation	Leverages an LLM to convert user intents, expressed in natural language, into structured search queries compatible with the 6G Data Space service catalogue.
REQ 6 Gaia-X Trust Compliance	Ensures all data operations adhere to Gaia-X trust, identity, and security frameworks, including authentication, authorization, and policy enforcement.
REQ 7 Curated Data set Availability	Guarantees access to a baseline of high-quality, pre-processed data sets originating from the project's testbeds, ready for reuse or further transformation.
REQ 8 Auto-Triggered ELT Workflows	Automatically initiates ELT pipelines upon data set retrieval or generation, streamlining data preparation tasks.
REQ 9 Data Cleaning & Augmentation Support	Integrates mechanisms for cleansing and enriching data sets through domain-specific algorithms, improving data set quality and model readiness.

REQ 10 Experiment Result Reintegration	Ensures that data sets generated from on-demand experiments are processed and added back into the 6G Data Space for future use by other researchers.
REQ 11 Natural Language to Experiment Translation	Uses an LLM trained to translate high-level user data needs into executable experiment plans, enabling automated testbed utilization.
REQ 12 Multi-Testbed Orchestration	Manages the selection and coordination of multiple 6G testbeds to run distributed experiments efficiently and in parallel, based on user requirements.
REQ 13 Testbed Access Control	Provides secure, programmatic access to the 6G testbeds, enabling controlled experiment deployment, resource allocation, and data collection.
REQ 14 Experiment Monitoring & Logging	Tracks real-time experiment execution across testbeds and logs metadata, status, and output for auditability and debugging.
REQ 15 Quality Assessment Metrics	Defines and applies performance and quality indicators to assess the value of data sets, including improvements from cleaning and augmentation.

4.5.1 Personae profiles and user stories for each PoC

This section focuses on the creation of the fictitious characters, by providing a detailed description of their roles and requirements' needs (Table 7, Table 12, Table 16). Additionally, a user story is elucidated by identifying its: (i) goal, (ii) acceptance criteria and (iii) general requirements. At the end of this exercise, each User story is mapped to a fictitious character. PoC#1, PoC#2 and PoC#3 user stories are described in Tables 8-11, 13-15 and 17-20, respectively.

4.5.1.1 Stories Creation for PoC#1

Table 7: Personae profiles for PoC#1

Persona	Description
Sofia Researcher (Academia / Industry)	<p>Sofia is a researcher working in the academic or industrial domain, focusing on the development of AI/ML models for next-generation communication systems. She needs access to reproducible, high-quality data sets and controlled experimentation environments to validate her hypotheses. Sofia is particularly interested in data sets that reflect real-world 6G conditions and infrastructure performance. She prefers intuitive tools that allow her to discover or generate data sets on demand and integrate them seamlessly into her model training workflows.</p> <p>Sofia as a researcher focused on developing AI/ML models for 6G communication systems, thus requires easy access to high-quality, well-labelled data sets and reproducible experimentation environments to test theoretical models and validate hypotheses. (User Story 1- Table 8)</p>
Mark (AI/ML Developer)	<p>Mark is a machine learning engineer responsible for training and fine-tuning AI models for 6G applications. He works with real-time data and has a high demand for well-labelled, cleaned, and contextualized data sets. Mark seeks environments that support hyperparameter tuning, transfer learning, and automated data processing through integrated ELT pipelines. He values natural language interfaces that reduce the overhead of manual query formulation.</p> <p>Mark as an AI/ML developer, wants to generate new data sets via on-demand experiments when existing data does not fit my needs so I can fine-tune my models with highly relevant inputs. (User Story 2 - Table 9)</p>
Elena (Data Platform as a Service Provider Representative)	<p>Elena is a cloud infrastructure provider managing services for hosting data platforms and AI pipelines. Her focus is on enabling scalable and interoperable execution environments across cloud and edge deployments. She provides infrastructure for data storage, model training, and ELT workflows. Elena prioritizes seamless integration with external data sources and support for trusted, compliant data exchange aligned with Gaia-X and IDSA standards.</p>

	Elena as a provider of cloud infrastructure and services that host AI models, data sets, and processing workflows focuses on offering scalable, secure platforms that can integrate easily with Data Spaces and experiment pipelines. (User Story 3- Table 10)
David (Data Provider)	David manages a 6G testbed. His role is to ensure that his infrastructure is ready, reliable, and utilized effectively for on-demand experiments and data set generation. He is responsible for exposing experimentation capabilities to external users via APIs and supporting orchestration, monitoring, and data export mechanisms. David wants to demonstrate the value and capabilities of his testbed within the broader 6G-DALI ecosystem. David as the Owner/operator of testbed infrastructure used in 6G experiments is responsible for ensuring resource availability, orchestrating experiments, and sharing usable data sets with the ecosystem. (User Story 4- Table 11)

Experiment 1.1 - Data search and extraction via Gaia-X service catalogue

Table 8: User Story 1 - Data set Discovery Using Natural Language in PoC#1

Goal	Enable Sofia to discover suitable data sets in the 6G Data Space using natural language interfaces.
Acceptance Criteria	<ul style="list-style-type: none"> • User provides a data set request via the intent-based Northbound API. • The system translates the request using LLM into a structured query. • The query is executed against the Gaia-X compliant catalogue. • Matching data sets are presented to the user within 5 seconds. • The user is informed if no suitable data set is found.
Relevant Requirements	REQ 4, REQ 5, REQ 6, REQ 7

Table 9: User Story 2 - Triggering ETL Pipeline for Existing Data sets in PoC#1

Goal	Allow Mark to automatically clean and augment discovered data sets before use.
Acceptance Criteria	<ul style="list-style-type: none"> • User selects a data set from the discovered results. • The system prompts to apply transformation. • Upon confirmation, an ELT pipeline is triggered. • Data cleaning and augmentation is applied using 6G-DALI algorithms. • Transformed data set is made available for download or further use.
Relevant Requirements	REQ 8, REQ 9, REQ 10

Experiment 1.2 - Data on-demand and enhanced ELT pipelines

Table 10: User Story 3 - Generating New Data set via Experiment Orchestration in PoC#1

Goal	Support Elena in generating data sets when no existing data set meets their needs.
Acceptance Criteria	<ul style="list-style-type: none"> • User submits a data request via the Northbound API. • The system confirms no suitable data set exists. • An LLM translates the request into an experiment plan. • The testbed selection mechanism chooses at least two testbeds. • Experiments are deployed on selected testbeds. • Raw data is collected and stored in staging.
Relevant Requirements	REQ 11,REQ 12, REQ 13, REQ 14

Table 11: User Story 4 - ELT Pipeline Execution for Generated Data in PoC#1

Goal	Ensure David that generated raw data is cleaned, augmented, and published back into the Data Space.
-------------	--

Acceptance Criteria	<ul style="list-style-type: none"> • Upon completion of experiments, raw data is passed to ELT pipeline. • Data is cleaned and augmented per predefined quality criteria. • A quality improvement of at least 30% is reported compared to raw data. • Data set is published in the 6G Data Space and catalogued.
Relevant Requirements	REQ 4, REQ 6, REQ 8, REQ 15

4.5.1.2 Stories Creation for PoC#2

Table 12: Personae profiles for PoC#2

Lucia (Vertical Service Provider)	Lucia works for a media company that delivers video streaming services across distributed edge networks. She needs to ensure optimal resource allocation and performance for video delivery by selecting the best ML model for resource prediction. Using the 6G-DALI framework, she benchmarks various pre-trained models available in the ML catalogue using data she owns, data available in the 6G-DALI Data Space, or data that are generated on demand. Her goal is to identify the model that delivers the highest prediction accuracy and robustness and best fits her application. (User Story 5- Table 13)
Mario (AI/ML Developer for Edge Applications)	Mario is an AI developer working on computer vision models for low-power IoT surveillance cameras operating at the network edge. His challenge is to fit high-performance models within limited memory and compute budgets without compromising accuracy. He uses the 6G-DALI framework to compress models (either he owns or from the ML catalogue) using configurable options such as the compression technique, the compression level, the validation strategy, and the evaluation metric. This enables him to deploy lightweight models at the Deep Edge nodes, maintaining efficiency and performance in real-world conditions. (User Story 6- Table 14)
George (Vertical Service Provider)	George manages smart city AI services that rely on predictive maintenance and anomaly detection models deployed across a distributed sensor network. He is focused on ensuring long-term reliability under varying real-world conditions. Using the 6G-DALI framework, he configures continuous performance monitoring and drift detection for the deployed models by selecting appropriate metrics and monitoring frequency. When drift is detected or model performance drops below a threshold, the system triggers retraining. George can retrieve updated models or publish new versions to the ML catalogue, ensuring continuity and trustworthiness. (User Story 7- Table 15)

Experiment 2.1 - Hyperparameter optimization and placement of ML models at the Cloud-Edge-Continuum

Table 13: User Story 5 - Model Compression in PoC#2

Goal	Allow Mario to compress ML models so that he can deploy them on resource-constrained devices, while preserving the required performance levels.
Acceptance Criteria	<ul style="list-style-type: none"> • The user can select the ML model (custom or available in the ML Catalogue) • The platform allows configuration of: <ul style="list-style-type: none"> ○ Compression technique ○ Compression level ○ Validation strategy ○ Evaluation Metric • The system returns a compressed model suitable for deployment on a specific target device. • The compressed model is available for download by the user.
Relevant Requirements	REQ2

Experiment 2.2 - ML model Benchmarking and drift detection during vertical application testing

Table 14: User Story 6 - Model Benchmarking in PoC#2

Goal	Allow Lucia to evaluate and compare pre-trained ML models from the 6G-DALI ML Catalogue.
Acceptance Criteria	<ul style="list-style-type: none"> • The user can select multiple models from the ML catalogue for benchmarking. • The benchmarking process supports the use of user-provided data sets, data sets available in the Data Space and data sets generated on demand. • Performance comparison metrics are provided for each model. • The best-performing model is identified based on user-defined criteria. • The selected model can be downloaded by the vertical.
Relevant Requirements	REQ1, REQ4, REQ10

Table 15: User Story 7- Model Drift Detection / Continuous Monitoring in PoC#2

Goal	Allow George to deploy and continuously monitor ML models and trigger retraining in case of performance degradation.
Acceptance Criteria	<ul style="list-style-type: none"> • The user can select the deployed model (custom or from the ML Catalogue). • The user can configure the drift metric and the detection frequency. • The system monitors the models and flags drift when the performance degrades, or statistical change is detected. • In case of drift detection, an automated retraining process is triggered using the latest data and a new model version is produced and made available for deployment or stored in the ML Catalogue. • The updated model meets predefined baseline performance or improvement threshold.
Relevant Requirements	REQ3

4.5.1.3 Stories Creation for PoC#3

Table 16: Personae profiles for PoC#3

Role	User Story
Alex (AI Expert & ML Developer)	Alex specializes in developing reinforcement learning agents for next-generation 6G networks. For PoC#3, he leverages the 6G-DALI framework and DTT to conduct large-scale experiments involving thousands of simulated mobile UEs and RAN configurations. His main challenge is to validate RL agents quickly and accurately across diverse traffic and mobility patterns. Using the automated experimentation pipeline provided by DTT and integrated AI framework components (user gateway, data cleaning, adaptation layer, DTT adapter), Alex can efficiently generate and use large data sets for model training. His goal is to reduce the RL validation time by 50%, ensure 80% automation, and generate data sets that generalize well from simulation to real-world deployment. (User Story 8- Table 17)
Sara (Data Provider)	Sara is responsible for managing the OEM DTT infrastructure that supports experimentation across both virtual and physical environments. She ensures that the testbed can handle both low-fidelity, large-scale simulations and high-fidelity, medium-scale trials using real radios and UEs. For PoC#3, Sara configures the network layout, device behaviour, and traffic patterns via the DTT's configuration manager. She enables deployment of xApps on RIC and supports NWDAF-based feedback loops. Her objective is to validate O-RAN scenarios with real signals and ensure that 100% of the generated data is integrated with the 6G Data Space and compliant with O-RAN standards. This strengthens the relevance and visibility of her testbed infrastructure. (User Story 9- Table 18, User Story 10- Table 19 and User Story 11- Table 20)

Experiment 3.1 - Large-scale experimentation for RLOps

Table 17: User Story 8 - Automated Reinforcement Learning Agent Validation in PoC#3

Goal	Alex wants to reduce the time required to validate RL agents by using large-scale simulations in the Digital Twin Testbed (DTT), leveraging the AI framework for automation.
Acceptance Criteria	<ul style="list-style-type: none"> • The user defines the RL agent to be validated. • The system sets up a large-scale virtual network scenario (multiple UEs, mobility, traffic) using VIA and OAI tools. • The experiment is executed automatically with no human in the loop. • Performance data (e.g., mobility success rate, handover delay) is collected and evaluated. • The RL agent is validated against defined KPIs. • Results and logs are stored and accessible via the 6G Data Space.
Relevant Requirements	REQ2, REQ12

Table 18: User Story 9 - Data set generation for Mobility Optimisation in PoC#3

Goal	Sara opts to generate diverse, high-volume data sets capturing UE mobility and network behaviour under different configurations to train RL agents.
Acceptance Criteria	<ul style="list-style-type: none"> • The user selects target KPIs (e.g., attach request success rate, throughput). • The DTT runs simulations involving high UE density and dynamic mobility. • Network scenarios include multiple gNBs and CN setups. • Data sets are automatically cleaned, labelled, and stored. • At least two distinct data sets are generated, supporting different mobility profiles. <ul style="list-style-type: none"> • Data is shared with stakeholders via the 6G Data Space.

Relevant Requirements	REQ7, REQ8, REQ10
------------------------------	-------------------

Experiment 3.2 - Medium-scale DT experimentation for O-RAN

Table 19: User Story 10 - Real-Time Monitoring and Adaptive Experimentation in PoC#3

Goal	Sara wants to enable real-time monitoring and dynamic adjustment of network parameters during O-RAN tests to optimize performance and collect targeted insights.
Acceptance Criteria	<ul style="list-style-type: none"> • The DTT control layer continuously monitors selected KPIs. • Custom xApps are deployed to NWDAF to trigger parameter adjustments during runtime. • Experiments adapt automatically based on current network conditions. • The entire process is logged and versioned for traceability. • Final results are validated against pre-defined thresholds. • All insights are stored in the 6G Data Space for reuse.
Relevant Requirements	REQ3, REQ4, REQ5

Table 20: User Story 11 - High-Fidelity O-RAN Scenario Testing in PoC#3

Goal	Sara aims to validate the performance of O-RAN-compliant network components using real radio hardware in realistic medium-scale scenarios.
Acceptance Criteria	<ul style="list-style-type: none"> • The user defines a network experiment with real UEs and radio equipment. • The DTT configures and deploys the scenario using VIA hardware. • The system captures KPIs such as handover latency, UE connectivity, and network throughput. • The experiment runs autonomously, without manual intervention. • All generated data is compliant with O-RAN standards. • Data is exported and shared via the 6G Data Space.
Relevant Requirements	REQ11, REQ12

5 Conclusions

This deliverable focuses on the business approach and use-cases assessment for the 6G-DALI project, which aims to address the crucial need for an intelligent, end-to-end AI framework to support the evolution of 6G Networks. The two pillars of 6G-DALI are; (i) Data and Analytics collection and storage via DataOps and (ii) AI Experimentation as a Service via Machine Learning Operations MLOps.

Furthermore, 6G-DALI focuses on three PoCs, those being; (i) *PoC#1-Data management and experiment on demand*, addressing the need for high-quality data sets, (ii) *PoC#2- AIaaS for CDN apps via cross-tested centralised MLOps*, addressing the decentralised MLOps capabilities of 6G-DALI and (iii) *PoC#3 – DTT and RLOps for large and medium scale experiments*, showcasing 6G-DALI’s ability to conduct extensive experiments through its AI framework and digital twin infrastructure. Under each PoC the associated KPIs and KVI are identified. Furthermore, a sustainability analysis of the e2e AI framework is performed based on the 3 PoCs outcomes.

The identification of the User requirements for each PoC, is performed through the formation of fictitious Characters’ User Stories, according to which virtual characters are created with specific needs to perform tasks and given roles such as AI/ML Developers or Cloud Infrastructure provider. A general list of requirements was initially formed after a discussion with the PoC leaders, and the personae profiles subsequently selected their requirements from the general bucket list.

The monetisation opportunities of 6G-DALI are thoroughly investigated, highlighting the reasons of “Data Scarcity” of telecom data sets and the strategic opportunity this creates for the project. 6G-DALI’s Stakeholders are identified as well as with the value proposition the project creates for them. The Gaia-X business model is also explained in this document because it directly informs and shapes the 6G-DALI business model. It is by adopting the Gaia-X’s principles that 6G-DALI framework creates a trusted dataspace where diverse stakeholders, as already identified, can effectively collaborate and innovate. In order to articulate the commercial and strategic potential of the 6G-DALI project, a BMC was chosen to be developed as a strategic tool.

6G-DALI aims to significantly impact the European technology landscape, fostering a competitive European Data Economy. Through its strategic alignment with major European 6G Initiatives, it serves as empowering stakeholders to innovate with confidence adhering to the European Values.

6 References

- [1] C. Mertens, "Establishing a Unified, Sovereign, and Open Digital Infrastructure: A Vision for Telecommunication Providers", *International Data Spaces Association*, March 2025, doi: 10.5281/zenodo.15041420.
- [2] M. D. Wilkinson, et al. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific data* 3.1 (2016): 1-9, doi: 10.1038/s41597-019-0009-6.
- [3] Huawei White Paper – Data Storage 2030. Available at: https://www-file.huawei.com/-/media/corp2020/pdf/giv/2024/data_storage_whitepaper_2030_en.pdf
- [4] Jones & Tonetti, AER (2020). Nonrivalry and the Economics of Data – Available at: <https://www.aeaweb.org/articles?id=10.1257/aer.20191330>
- [5] <https://edgedelta.com/company/blog/data-market-size-and-forecast>
- [6] <https://www.marketsandmarkets.com/Market-Reports/6g-market-213693378.html>
- [7] <https://www.avenga.com/magazine/how-to-generate-value-with-data-monetization-in-telecom/>
- [8] <https://www.marketsandmarkets.com/Market-Reports/artificial-intelligence-ai-as-a-service-market-121842268.html>
- [9] <https://aibusiness.com/generative-ai/domain-specific-ai-models-poised-to-dominate-enterprise-gartner>
- [10] <https://5g-ppp.eu/wp-content/uploads/2017/07/5G-PPP-5G-Architecture-White-Paper-2-Summer-2017-For-Public-Consultation.pdf>
- [11] <https://www.techuk.org/resource/five-barriers-to-managing-legacy-systems-for-the-purposes-of-data-sharing.html>
- [12] Mohamed Mekki, Nassima Toumi, & Adlen Ksentini, "Benchmarking on Microservices Configurations and the Impact on the Performance in Cloud Native Environments (1.0).", *LCN 2022, 47th Annual IEEE Conference on Local Computer Networks*, Edmonton, Canada. Zenodo, doi: 10.5281/zenodo.6907619.
- [13] A. Colpaert, S. De Bast, R. Beerten, A. P. Guevara, Z. Cui and S. Pollin, "Massive MIMO Channel Measurement Data Set for Localization and Communication," in *IEEE Communications Magazine*, vol. 61, no. 9, pp. 114-120, September 2023, doi: 10.1109/MCOM.004.2200716.
- [14] A. Colpaert, Z. Cui, E. Vinogradov and S. Pollin, "3D Non-Stationary Channel Measurement and Analysis for MaMIMO-UAV Communications," in *IEEE Transactions on Vehicular Technology*, vol. 73, no. 5, pp. 6061-6072, May 2024, doi: 10.1109/TVT.2023.3340447.
- [15] M.Mekki, N. Toumi, A. Ksentini, "Benchmarking on Microservices Configurations and the Impact on the Performance in Cloud Native Environments", in *LCN 2022, 47th Annual IEEE Conference on Local Computer Networks*, August 2022, doi: 10.5281/zenodo.6907618.