This project has received funding from the European Union's Horizon Europe research and innovation program under grant agreement No 101120823

MANiBOT

Advancing the physical intelligence and performance of roBOTs towards human-like bi-manual objects MANipulation

D7.1. System testing and demonstration plan

WP number and title	WP7 – MANiBOT framework testing, demonstration and validation
Lead Beneficiary	Fraport Greece (FG)
Contributor(s)	CERTH, SSSA, UBU, TUW,THL,CIOP, SDI, Masoutis
Deliverable type	Report
Planned delivery date	31/10/24
Last Update	30/10/24
Dissemination level	PU - Public

Disclaimer

This document reflects only the author's view. Responsibility for the information and views expressed therein lies entirely with the authors. The European Commission are not responsible for any use that may be made of the information it contains. The MANiBOT Consortium consists of the following partners:

Participant No	Participant organisation name	Short Name	Country
1	ETHNIKO KENTRO EREVNAS KAI TECHNOLOGIKIS ANAPTYXIS	CERTH	GR
2	ARISTOTELIO PANEPISTIMIO THESSALONIKIS	AUTH	GR
3	TECHNISCHE UNIVERSITAT DARMSTADT	TUDa	DE
4	SCUOLA SUPERIORE DI STUDI UNIVERSITARI E DI PERFEZIONAMENTO S ANNA	SSSA	IT
5	UNIVERSIDAD DE BURGOS	UBU	ES
6	TECHNISCHE UNIVERSITAET WIEN	TUW	AT
7	TWI ELLAS ASTIKI MI KERDOSKOPIKI ETAIREIA	THL	GR
8	CENTRALNY INSTYTUT OCHRONY PRACY - PANSTWOWY INSTYTUT BADAWCZY	CIOP	PL
9	ASEA BROWN BOVERI SA	ABB	ES
10	FRAPORT ETAIRIA DIACHEIRISIS TON PERIFEREIAKON AERODROMION TIS ELLADAS ANONYMI ETAIREIA	FG	GR
11	SCHWARZ DIGITAL GMBH & CO. KG	SDI	DE
12	DIAMANTIS MASOUTIS AE SUPER MARKET	MASOUTIS	GR
13	UNIVERSITY OF BRISTOL	UoB	UK

Document Authors

Partner	Author
FG	Anastasios Symeonidis, Angeliki Gkalimaridou, Ioanna Karountzou Margarita Kontolatou, Michel Stamoulis
Masoutis	Eleftheria Matta, Panagiotis Kalaidopoulos, Kyriakos Pantouvakis, Apostolos Doxanidis, Panagiotis Amoutzias
SDI	Valentin Dürlich, Jonathan Schatz
UBU	Jesús Enrique Sierra García, Mario Peñacoba Yagüe

Document History

Version	Date	Status	Description
0.1	03/07/24	Draft	Table of Contents
0.2	15/10/24	Draft	First draft completed
0.3	21/10/2024	Draft	Changes proposed by CERTH were incorporated
0.4	30/10/2024	Draft	Changes proposed by the reviewers (ABB, TUDa) were incorporated
1.0	31/10/2024	Final version	Quality check performed, final version ready to be submitted

Definitions, Acronyms and Abbreviations

Acronyms and Abbreviations	Description
BTRS	Baggage Tracking and Reconciliation System
DG RTD	Directorate-General for Research and Innovation
FIFO	First In First Out
GA	Grant Agreement
HE	Horizon Europe
KPI	Key Performance Indicator
UC	Use Case
UI	User Interface

Executive Summary

The present document is a deliverable of the MANiBOT project, funded by the European Commission's Directorate-General for Research and Innovation (DG RTD) under its Horizon Europe programme (HE).

The objective of this deliverable (D7.1) is to outline the strategy for the system testing and the demonstration plan of the bi-manual, mobile service robot developed under the MANiBOT project. Through the testing and demonstration, we aim to ensure the robot's readiness for deployment to all Use Cases (UCs), addressing both technical specifications as well as user requirements under the targeted Key Performance Indicators (KPIs).

This deliverable will guide the activities of Task T7.1, which provides a description of the pilot and lab facilities along with the planning of implementation, integration, testing, execution, and validation activities of pilots. The deliverable will first present the capabilities and functionality of each test facility. It will also outline the major milestones for modules integration testing and set the methodology to record the risks and issues related to the MANiBOT testing and evaluation. In the updated version of the deliverable (D7.5 to be submitted on M22) the final testing plans will be analysed including the plans of integration testing. The detailed UC scenarios to be tested on the pilot sites will be described and the demonstration plans will be presented.

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Scope of the deliverable

The scope of this deliverable is to ensure that the robot meets all functional, technical, and user requirements, with a particular focus on its application in baggage handling at airports and restocking supermarket shelves. This deliverable is critical to validating the effectiveness and reliability of the MANiBOT in performing a wide variety of manipulation tasks, with highly diverse objects, in a human-like manner and its performance, in diverse, challenging environments.

Objectives:

- Plan when the different capabilities of the robot will be tested during development
- Consider the prerequisites for testing implementation, defined in D7.2
- Consider the Use Case (UC) KPIs for testing implementation, defined in D2.2
- Plan how the robot will be demonstrated and evaluated
- Set a timeline for the testing activities
- Identify risks and issues for the MANiBOT system testing and demonstration

1.2 Relation to other Activities and Deliverables

This deliverable provides a description of the system testing and demonstration plan, taking into account the requirements, use cases and KPIs as defined by WP2. Specifically, the user requirements and use case scenarios (UC1-UC4), have been determined under T2.2 works and have been reviewed in D2.2 (M12). In M22 all updates in the use case scenarios based on the development and early deployment feedback, will be included in the final version of D7.1, namely, D7.5.

The project's key objectives and their respective KPIs have been described in the Grant Agreement (GA) document and have been revised under T2.2, related to D2.2, after the UCs finalization.The robot functionalities, technical specifications, and overall system architecture of MANIBOT, along with related technical KPIs, have been defined in D2.4, which is connected to T2.4. Taking in account that the components to be developed by WP3, WP4, WP5 and WP6 will be tested and evaluated according to the established plans, this deliverable also relates to all these technical tasks.

Additionally, ethics and data management tasks (WP9, T1.4), will oversee the tests and evaluation processes, particularly at the pilot sites, to ensure that all activities comply with relevant regulations and standards.

This deliverable later during the project will be the basis to produce D7.5, while it will be the guide for implementing T7.3, T7.4 and T7.5 activities.

1.3 Structure of the deliverable

The deliverable is structured as reported below:

Chapter 1 – Introduction – Provides a summary of the deliverable.

Chapter 2 – Testing Sites – Provides brief descriptions of the testing sites.

Chapter 3 – MANiBOT Use Cases and Components – Provides an overview of the functional components and use cases as these have been defined up to the time this deliverable has been prepared.

Chapter 4 – Testing and Validation – Provides the test framework, together with the approach to define the MANiBOT testing plan.

Chapter 5 – Description of validation criteria and KPIs – Provides the categories of KPIs used for MANiBOT project.

Chapter 6 – Risk & Issue Management Plan – Presents the risk & issue management methodology.

Chapter 7 – Conclusions – The deliverable concludes and describes future action.

2 Testing Sites

The purpose of this chapter is to provide a high-level description of the testing sites where the MANiBOT will be tested and demonstrated. A more detailed description is included in deliverable D7.2 (M12). Testing will occur in both lab environment and at pilot sites. In the lab environment, two different setups will be created to simulate airport-like and supermarket-like conditions, tailored to different use cases. Although the supermarket-related use cases (UC1 and UC2) have differences, a single supermarket-like setup will be sufficient in the lab environment because the core tasks of manipulating individual items and handling boxes can be tested under similar conditions, with any adjustments needed for specific tasks (more details can be found in D7.2). The pilot sites will focus on real-world conditions for the different use cases, ensuring the robot's performance is thoroughly validated in operational settings. The following sections outline both the lab and pilot sites environments correlated to each use case.

2.1 Lab Environment Setup

The lab environment for the MANiBOT project is located within the "La Milanera" technology park on the campus of the University of Burgos (Figure 1). This location was chosen due to its strategic proximity to the university's academic and research facilities.

Figure 1 Lab Location

The lab will accommodate two distinct setups reflecting the different use cases of the project. One setup will replicate the conveyor belt systems found in airport baggage handling areas, allowing researchers to develop robotic solutions tailored to automation challenges in airport logistics. The second setup will simulate a supermarket environment, enabling the study of robotics applications in retail scenarios. In particular, the areas designated for the supermarket and airport laboratory zones are highlighted in Figure 2.

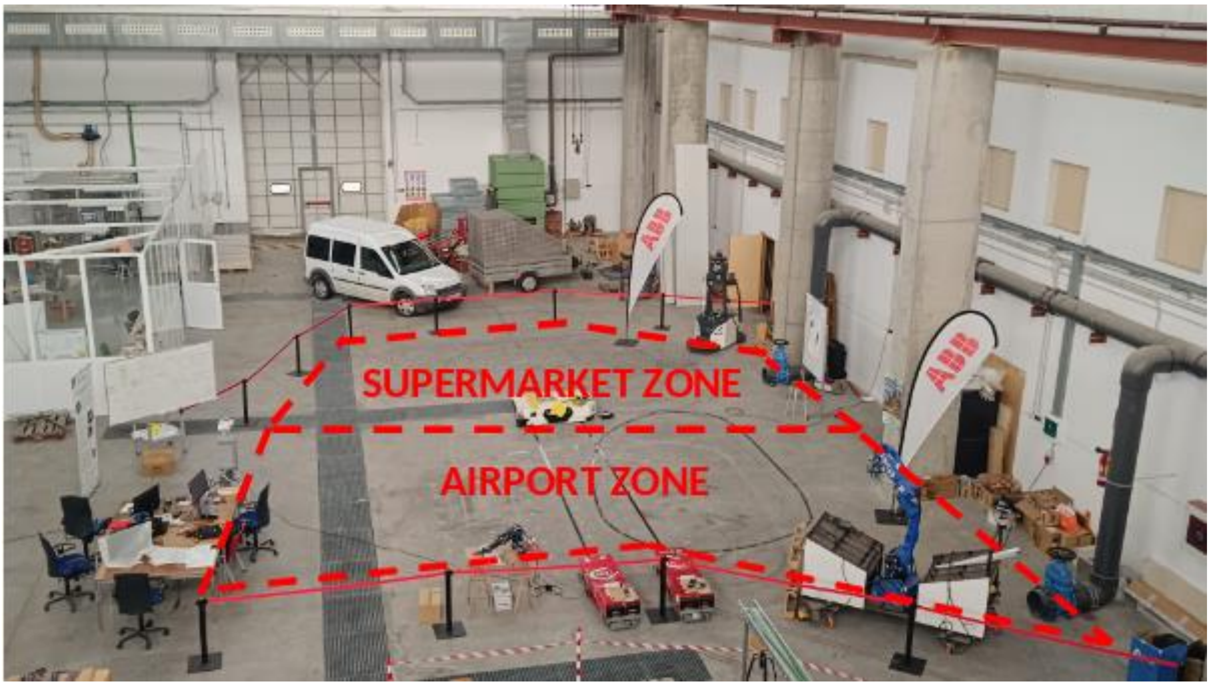

Figure 2 Supermarket & Airport Zones

2.2 Pilot Sites Setup

2.2.1 Use Case 1 - Single item manipulation for supermarket shelves restocking

The pilot site for Use Case 1 will be provided by partner Masoutis. It is the operational store "Grand Masoutis" located in Thermi, Thessaloniki, Greece (Figure 3). It was selected due to its large-scale operations, standard store layout, and extensive customer traffic, making it an ideal environment for testing the robot's efficiency and adaptability. This store's layout aligns with other large stores in the chain, ensuring that training the robot there will allow for seamless integration across similar locations. The variety of products and the spacious floor plan further enable comprehensive testing with minimal disruption to customers, while proximity to company headquarters ensures easy access for the project team.

Figure 3 Grand Masoutis in Thermi, Thessaloniki

2.2.2 Use Case 2 - Boxes of items manipulation during shelves restocking

The pilot site for Use Case 2 will be provided by partner Schwartz Group. It is one of Schwarz Group’s designated test stores located in Heilbronn, Germany (Figure 4). The test store is equipped with representative shelves filled with real or photorealistic dummy products (replicas), supporting the test of various product groups and can be used as well for early-stage testing of new technologies. The store replicates the layout of retail branches making it an ideal site for testing and demonstrating the MANiBOT robot.

Figure 4 Representative shelves within Swartz Group Pilot Site in Heilbronn, Germany

2.2.3 Use Case 3 & 4 - Loading baggage from conveyor belt to cart & vice versa, from carts to conveyor belts

The pilot site for UC3 and 4 will be provided by partner Fraport Greece (FG). It is the operational environment of Thessaloniki Airport “Makedonia” (SKG) located in Thessaloniki, Greece. For operational safety reasons of the airport as well as constraints imposed by the airport’s users, specific locations of the baggage handling area will be separated in order to test MANiBOT in parallel with live operations. Specifically, for UC3, a conveyor belt to the baggage sorting area will be used, while for UC4 an arrivals intake conveyor belt will be selected.

Figure 5 Baggage sorting area in Thessaloniki Airport “Makedonia”, Greece

Figure 6 Arrivals intake conveyor belt in Thessaloniki Airport “Makedonia”, Greece

3 MANiBOT Use Cases and Components

The purpose of this chapter is to provide an overview of: 1) the MANiBOT system components and, 2) the MANiBOT use cases, both of which constitute the fundamental items for the testing process. The system components are vital as they will be tested to ensure optimal development of the robot while the use cases are important since their demonstration by the robot will verify that it meets both technical and functional requirements. The UC and sub-use case description from the users' perspective are analysed in D2.2 and the description of the components in D2.4. The final UC scenarios to be demonstrated in the pilots will be described in D7.5.

3.1 MANiBOT Use Cases

The use cases selected for testing and evaluation are related with the retail and transport. These use cases have been chosen to provide challenging real-world settings that will significantly impact the proposed solutions, once adopted in the future. The focus is on four complex tasks that are performed daily by millions of people worldwide. These tasks include the manipulation of commercial items in supermarkets and the handling of baggage items in airports and are:

- The manipulation of the various commercial items sold in supermarkets [UC1] and boxes thereof [UC2], particularly focusing on the task of shelves restocking.
- The handling of baggage items in airports, focusing on the tasks of baggage loading and unloading from conveyor belts to carts [UC3] and vice versa, from carts to conveyor belts [UC4].

3.1.1 Use Case 1: Single item manipulation for supermarket shelves restocking

This UC involves the manipulation of individual commercial items sold in supermarkets, particularly focusing on the task of restocking shelves. The use case describes the process by which the MANiBOT robot autonomously performs product replenishment in retail stores. It includes the robot's ability to navigate the store, identify products and their correct shelf locations, restock shelves following First In First Out (FIFO) principles, and handle surplus or damaged products. Each step emphasizes the robot's interaction with its operator through the User Interface (UI), customers, and store employees, as well as challenges such as misidentifying products or encountering obstacles. The goal is for the robot to ensure accurate product placement, efficient task completion, and seamless integration with store operations.

3.1.2 Use Case 2: Boxes of items manipulation during shelves restocking

This UC deals with the manipulation of boxes containing multiple commercial items, focusing on the task of restocking supermarket shelves. The robot begins by navigating to the defined position, i.e. next to the mixed pallet with boxes of products, identifying the targeted product, and preparing them for replenishment. It applies the FIFO principle to ensure the proper product replenishment, overstock management and that products are placed correctly. Once all tasks are completed, the robot returns to its charging station. The process involves interaction with the UI and the presence of customers and employees, with potential challenges like blocked paths or unreadable labels. The goal is to ensure accurate and efficient restocking while maintaining high success rates for product recognition and placement.

3.1.3 Use Case 3: Loading objects from conveyor belt to cart

This UC focuses on the handling of baggage items in airports, specifically the tasks of autonomously loading baggage from conveyor belts to carts. The process starts with the operator assigning a task, and the robot positioning itself at the optimal location between the conveyor belt and baggage cart. The robot then recognizes and identifies baggage and reads bag tags to determine the flight, destination, and weight. Once identified baggage relate to its task, the robot transfers and organizes the baggage onto the cart, recording and transmitting bag tag information to the airport's BTRS system. After completing the loading process, it rests or returns to the charging position. The process involves challenges like unreadable bag tags, cluttered baggage, and potential obstacles in the working area, while safety and task accuracy are the key priorities.

3.1.4 Use Case 4: Unloading baggage from cart to loading belt

This UC focuses on the autonomous unloading of baggage from carts to conveyor belts by MANiBOT. The process begins with the robot moving from its resting or charging position to the designated area next to the arrivals belt. Once the baggage cart arrives, the robot ensures the area is clear of obstacles and positions itself optimally for unloading. The robot then identifies, manipulates, and transfers baggage onto the conveyor belt, reads bag tags and ensures their proper placement. After unloading each cart, the robot moves to the next cart, rests or returns to the charging position. Key challenges include detecting obstacles, handling improperly stacked and tagged baggage, and ensuring smooth coordination with human workers.

3.2 MANiBOT Components

The MANiBOT is composed of several key components, each designed to meet specific technical and performance requirements.

Based on the D2.4, 19 preliminary functional components have been identified, which are depicted in Figure 7 together with interconnections between them. These 19 functional components are mapped to specific technical tasks essential for the system's operation and are further broken down into 46 sub-components, each contributing to the detailed functionality and performance of the MANiBOT system. In Table 1 all the identified subcomponents can be found.

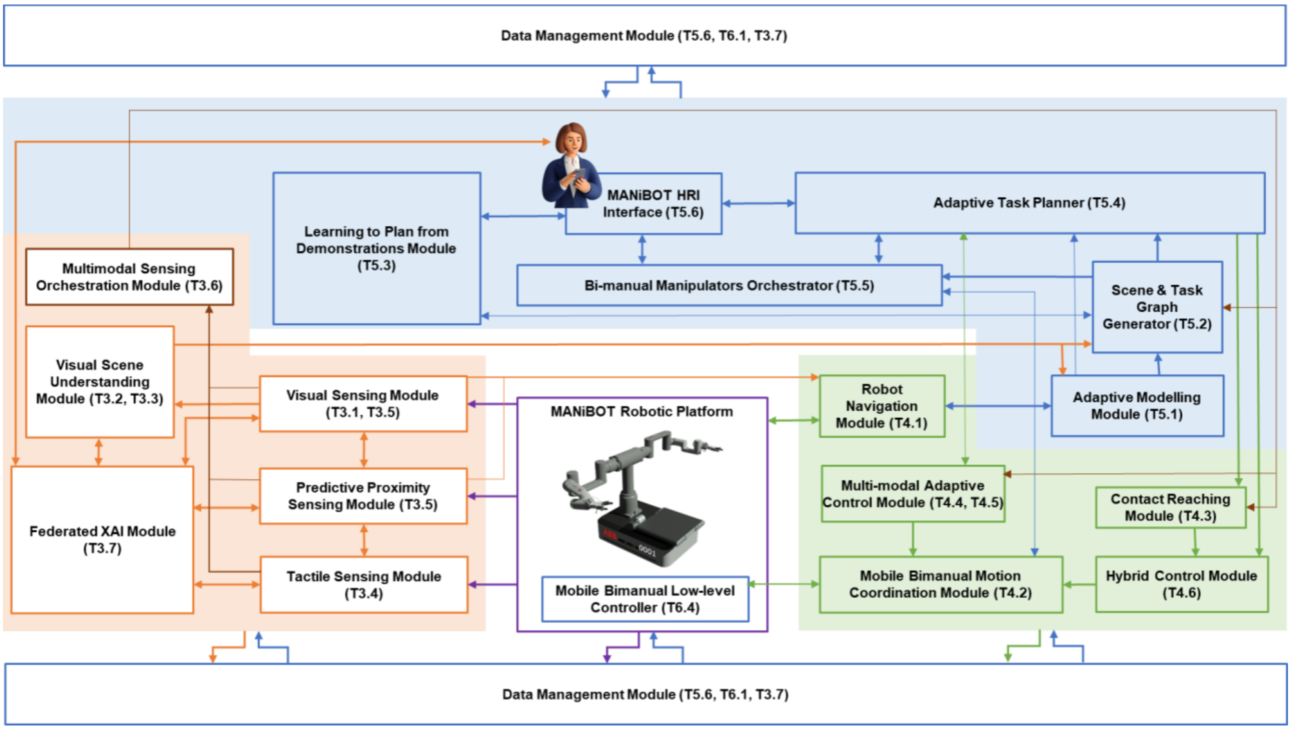

Figure 7 MANiBOT Functional Components & interconnections

Table 1 MANiBOT Functional Components

No	Functional Components	Sub-components	Related WP	Responsible partner
1	Visual Sensing Module	• Object Detection Module (T3.1) • 6DoF Object Pose Estimation Module (T3.1) • Human Pose Estimation and Tracking Module (T3.5)	WP3	CERTH/TUW
2	Tactile Sensing Module	• Contact Pose and Force Estimation Module (T3.4) • Contact Detection Module (T3.4)	WP3	UoB
3	Predictive Proximity Sensing Module	• Human Detection Module (T3.5) • Human Intention Predictor (T3.5)	WP3	SSSA/CERTH
4	Multimodal Sensing Orchestration Module	• Robot Platform Proprioceptive Data Analyser (T3.6) • Robot Platform Exteroceptive Data Analyser (T3.6) • Multimodal Data Orchestrator (T3.6)	WP3	CERTH
5	Visual Scene Understanding Module	• Target Object Semantic Segmentation Module (T3.2) • Predictive Structural Inference Engine (T3.3)	WP3	CERTH
6	Federated XAI Module	• Federated Learning Module (T3.7) • Explainable AI (XAI) Module (T3.7)	WP3	THL
7	Mobile bimanual low-level controller	• Motion Control Module (T4.2) • Bimanual Control Module (T4.2) • Conveyor Belt Control Module (T4.2)	WP4	AUTH
8	Robot Navigation Module	• Localisation Module (T4.1) • Global Planner (T4.1)	WP4	CERTH

		• Local Planner (T4.1)		
9	Multi-modal adaptive control Module	• Bimanual Manipulation Primitives Controller (T4.4-T4.5) • Unimanual Manipulation Primitives Controller (T4.4-T4.5) • Grasping Controller (T4.5)	WP4	AUTH
10	Mobile Bimanual Coordination Module	• Dual Arm and Mobile Platform Coordinator (T4.2) • Performance Optimisation Module (T4.2) • Collision Avoidance Submodule (T4.2)	WP4	AUTH
11	Contact Reaching Module	• Contact Selection Planner (T4.3) • Contact Reaching Controller (T4.3)	WP4	AUTH
12	Hybrid Control Module	• Hybrid controller (T4.6)	WP4	AUTH
13	Adaptive Modelling Module	• Semantic 3D SLAM module (T5.1) • Operational Environments Adaptable Model Generator (T5.1) • Lifelong Semantic Adaptation Module (T5.1)	WP5	CERTH
14	Scene & Task Graph Generator	• Scene Graph Generator (T5.2) • Task Graph Generator (T5.2) • Grounding Module (T5.2) • Primitive Sequencing Module (T5.2)	WP5	TUDa
15	Learning to Plan from Demonstrations Module	• Hybrid RL High-level Planner (T5.3) • Low-level Primitives Adapter (T5.3)	WP5	TUDa
16	Adaptive Task Planner	• Manipulation Task Assessment & Adaptation Module (T5.4) • Recovery Strategies Enforcer (T5.4)	WP5	TUDa

17	Bi-manual Manipulations Orchestrator	• High-level Task Orchestrator (T5.5) • Natural Language Feedback Integrator (T5.5)	WP5	TUDa
18	MANiBOT HRI Interface	• HRI GUI (T5.6) • Back-end server (T5.6) • AR Module (T5.6) • Task Schedule and Monitoring Module (T5.6)	WP5	CERTH
19	Data Management Module	• HRI Data Manager (T5.6) • Robotic Platform Data Manager (T6.1) • Federated Learning Data Manager (T3.7)	WP3, WP5, WP6	CERTH/THL

Based on the functional components presented above, as the project progresses, the detailed technical components will be defined that will constitute the testing items.

4 Testing and Validation

4.1 Testing and Validation Framework

The testing and validation framework provides the structured approach used to assess the functionality and performance of the MANiBOT system. This begins with the definition of the key concepts and continues with the methodology to be applied, including the different stages of testing which have been designed to verify specific aspects of the system, ensuring that all components function correctly individually and seamlessly integrate with one another. Additionally, this section highlights the objectives and milestones guiding the testing process, along with test cases content and review procedures to be followed.

4.1.1 Definitions

Test and Validation Framework is the approach that MANiBOT project will use for testing every unit of the system to ensure that it is operating as it should and that it properly interacts with the rest of MANiBOT components. It will also test and valiadate the overall system and ensure that it is performing in accordance with the system functional requirements and technical specifications with focus on those on performance.

Unit is a basic testing element that has a specified function, and which will be one of the forming parts of the system.

Module is a combination of units that work together to perform a specific task.

Preliminary Prototype is the preliminary version of the robot that will be created by integrating all its components to test and validate its functionality, usability, and feasibility as a whole.

Final Prototype is the final version of the robot that will be derived from the preliminary prototype after performing appropriate improvements and refinements as these emerged from prototype testing.

Lab tests are the tests that will take place at the different set ups of the lab environment.

Field tests are the tests that will take place at the different pilot sites.

Test case is a document including the specifications of the inputs, execution conditions, testing procedure and expected results that define a single test to be executed to achieve a particular objective.

4.1.2 Methodology

The Test Framework for the MANiBOT project is designed to ensure a thorough validation of the MANiBOT system's functionality, reliability, and readiness for deployment. It includes four testing stages: unit tests, integration tests, system tests and validation, which will be performed in different locations, to ensure both step by step testing at various levels of the system development and under different conditions.

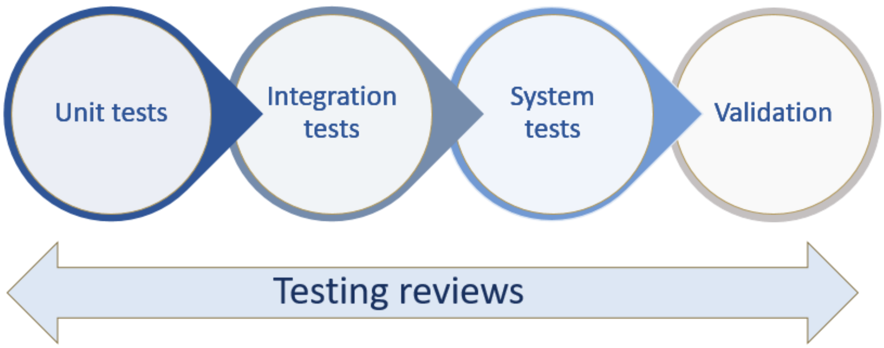

Figure 8 Testing Stages

- **Unit tests** will help to determine whether different individual components fulfill functionalities. Depending on the unit's technology and its requirements, the unit tests might be divided in two phases: 1) the simulation-based unit testing and, 2) the real-world unit testing. The necessity for testing in simulation environments (e.g. IsaacSIM [1], GaZeBo [2]) results either from the need of a big amount of data to be used both for training and initial testing, or because the units require HW that is not yet available (bought or developed).
- **Integration tests** will aid to verify interoperability between units. Similarly to the unit tests, integration tests might have two phases: 1) the simulation-based testing and, 2) the real-world testing. The simulation environments will be used for initial Inputs/Outputs (I/O) and connectivity communication testing between different units for early clarification of the units' interconnections and detection of misalignments in the expected communication between them. More technical details on the integration procedure and testing will be addressed in T6.5 and T6.6 and their corresponding deliverable D6.4 "Integrated bimanual mobile manipulator robot".
- **System tests** will help to examine the complete robot as a single system, starting from its preliminary prototype version towards the final one.
- **Validation** will verify that the requirements have been met .

Each test stage serves a unique purpose in the testing process, offering different levels of granularity and coverage, while the combination of all four test stages makes the MANiBOT test framework a balanced testing strategy which offers a comprehensive test coverage, early detection of errors and ensures the reliability and high-quality of the system to be developed.

Furthermore the decision to build at the lab environment two different setups that will simulate airport-like and supermarket-like conditions, tailored to different use cases, ensures that the solution to be developed will be both robust and flexible, with the potential to be applied across various real-world scenarios and will not only be conceptually strong, but will also be practically implementable. In addition, the performance of system tests and validation at the pilot sites before the evaluation will provide valuable feedback on the robot's behavior in actual operating environment.

It should be clarified that during the testing procedure intense efforts and focus will be gradually moved from the individual lab environments of technical partners, who will perform unit tests and several integration tests, to UBU labs, where the efforts will mostly focus on integration and system tests. The validation will be performed at the pilot sites based on the technical KPIs, before the final demonstration and evaluation of the system which will be based additionally to the use case KPIs.

Thus, this structured methodology ensures that all aspects of the MANiBOT will be thoroughly tested, from controlled lab settings to real-world pilot environments, resulting in a reliable and high-quality robot.

4.1.2.1 Unit Tests

Unit tests are the first level of the testing process and the most fundamental type of testing. Their objective is to test the correctness of individual components and illustrate that the component is accurately meeting the requirements, working according to specifications and delivering the expected output. These tests are crucial for identifying and resolving issues at an early stage. As they focus on individual components of the MANiBOT system, they are conducted at the facilities of the partner that develops the unit who is also responsible for the appropriate unit testing procedure. These tests are not included/described in this deliverable.

4.1.2.2 Integration Tests

Integration tests are the second level of the testing process during which different units that need to interact with each other will be combined to work as a group. Their objective is to test interactions between constituent units to verify the expected behavior of the module and to ensure that the combined units work together correctly as a group to achieve the desired functionality and performance. These tests are important, because they aid in the identification of integration issues between units which at this level are more straightforward to resolve than discovering them at the next testing level, system tests. In addition, this early detection prevents these issues from evolving into more complex problems during later stages of development. Integration tests will start with the units' availability and are expected to end once the MANiBOT preliminary prototype is available, ensuring that the different modules work seamlessly together. They will be conducted mainly at the lab environment of UBU as well as at the pilot sites where applicable.

4.1.2.3 System Tests

System tests are the third level of the testing process during which all components are tested together as a whole. For these tests, the preliminary prototype will be used, at various stages of its development. Their objective is to evaluate how the various components interact together in the integrated robot and to verify that the preliminary prototype performs tasks as designed in a real-world environment. These tests are critical, because they provide important feedback to improve the final prototype before proceeding to validation, whereas they also minimize the risks associated with the behavior of the robot in its operating environment. System tests will start once the preliminary prototype is available and will end with the availability of the final prototype, to make sure that the prototype meets specified requirements and runs as smoothly as possible in its operating environment. They will be conducted initially at the lab environment and at the pilot sites upon their readiness.

4.1.2.4 Validation

Validation is the last and final level of the testing process during which the robot will be further tested to confirm that it complies with the end-user requirements based on the technical KPIs as these described in the relevant Chapter. Their objective is to validate and verify that the robot not only functions as intended but also meets the needs and expectations of its users, providing confidence in its readiness for use. These tests determine whether the robot should be deployed at the pilot sites for demonstrations and evaluation, and they serve as a quality assurance checkpoint. Validation will start when all the feedback received by the system tests will have been implemented to the preliminary prototype in order to validate the robot's functionality, performance and usability in live environments. Validation will be implemented in two phases; preliminary and final validation. During the first phase, the preliminary prototype will be used, while at the second phase, the final prototype will be used. Validation will be conducted at all three pilot sites.

4.1.3 Objectives

The primary objectives of the tests to be performed are to ensure MANiBOT's:

- **Functionality**: This is a foundational test objective, which ensures the developed module/system functionalities as intended. During functional tests, the developed module/system will be evaluated against the specified requirements. For example, functional testing for the robot includes the tasks the robot must perform.
- **Performance**: Determine if the developed module/system meets specific quality requirements. MANiBOT will focus, among other performance criteria, on elements such as accuracy, speed and usability. Tests on accuracy will validate that MANiBOT performs tasks with a high degree of precision. Through speed tests the efficiency and responsiveness of the developed module/system will be assessed by measuring the time required to perform the different tasks. Evaluation on how the developed module/system interacts with users and responds to commands will be performed through usability tests during which it will be defined how easy it is to use the module/system and if there are any problems that need to be fixed.
- **Reliability**: Aim to confirm the consistency of the MANiBOT under different conditions, ensuring minimal downtime and robust performance. During reliability tests, the developed module/system is subjected to various stress conditions and circumstances over an extended period to identify potential failures or malfunctions.
- **User acceptance:** Aim to ensure that the MANiBOT will be trusted by the end users in terms of reliability, safety, transparency, technical competence, ethical behavior and emotional engagement. To achieve trustworthiness an assessment will be performed, within T2.3, which will be taken in account to improve human-robot interactions.
- **Safety**: Ensure that the MANiBOT operates safely around humans and in complex environments, adhering to all safety standards and protocols.

These objectives guide the testing process to ensure that the MANiBOT is ready for real-world deployment and capable of meeting the project's goals effectively. It should be noted that the testing process at the pilot sites will be closely observed in terms of ethical, legal and safety issues by the corresponding MANiBOT responsible (AUTH, T1.4).

4.1.4 Milestones

The main milestones posed by the project plan included in the Grant Agreement are depicted in the Table below.

Table 2 Grant Agreement Milestones

Milestone No	Milestone Name	Lead beneficiary	Due Date (month)
5	Preliminary integrated system available to lab testing	CERTH	August 2025 (M22)
7	Preliminary integrated system lab tests start	UBU	July 2026 (M33)
10	Final pilot sites preparation established	UBU	October 2026 (M36)
11	MANiBOT robot platform available to pilots	UBU	January 2027 (M39)
12	Project tests and demonstrations performed	FG	April 2027 (M42)

Taking these milestones into account and matching them with the testing methodology described, specific milestones for the testing stages have been extracted which are depicted in the following Table.

Table 3 Testing stages milestones

GA Milestone No	GA Milestone Name	Test Stage Milestone Name	Due Date (month)
5	Preliminary integrated system available to lab testing	Extensive integration tests on the preliminary versions of core MANiBOT modules focusing on connectivity issues, achieving main functionalities	August 2025 (M22)
7	Preliminary integrated system lab tests start	Start of system tests with the preliminary prototype demonstrating the use cases at lab environment using the testbed	July 2026 (M33)
10	Final pilot sites preparation established	Start of system tests at pilot sites	October 2026 (M36)
11	MANiBOT robot platform available to pilots	Start of final validation with the final prototype	January 2027 (M39)
12	Project tests and demonstrations performed	Final Evaluation (T7.5, D7.4)	April 2027 (M42)

Integration tests of the core MANiBOT components focusing on establishing basic functionalities will start the latest on August 2025. The integration will be continuously enriched by new and updated versions of the MANiBOT modules and thus iteratively tested. Once the preliminary prototype of the MANiBOT bi-manual robot will be released (July 2026) the system testing at the lab environment will begin demonstrating parts of the use-cases. Upon pilot sites readiness the latest in October 2026, system tests will be continued at the pilot sites until January 2027 that the last phase of development will finish, and the final prototype will be available for the final validation. Within this period, October 2026 – January 2027, also preliminary validation will be performed at the pilot sites with the use of the preliminary prototype.

During the course of the project and upon the testing plan development additional milestones will be defined, such as when the validation will take place at each pilot site. Thus, the final milestones for the testing process will be further elaborated and reviewed in subsequent phases.

4.1.5 Test Cases

Aiming to facilitate the testing process and to have a methodical approach, before starting the tests execution of each testing stage, test cases will be prepared for all the tests to be performed. The test cases will be also used during the tests execution in order to record the actual testing results. The test case template will include at least the below key components.

- **Test case number:** A unique identifier for the test case.
- **Test description:** A small description about what the test concerns.
- **Test scope:** What is included and excluded from the test.
- **Test items:** List with the units, modules, prototype or system and the features that will be tested.

- **Test objective:** The test goal in accordance with the test framework, e.g. "Functionality," "Performance," "Reliability", "User acceptance" and "Safety."
- **Test stage:** The testing stage in accordance with the test framework, i.e. "Unit," "Integration," "System," "Validation."
- **Test steps:** The exact steps to perform to execute the test.
- **Pre-requisites:** Preparation required before test execution.
- **Entry criteria:** Conditions that must be met before testing begins.
- **Exit criteria:** Conditions that must be satisfied to conclude testing.
- **Expected result:** The expected test outcome and criteria to determine whether it has been realized.
- **Risk identification:** List with potential risks that could impact the test.
- **Risk treatment:** Description of how the identified risks will be addressed.
- **Tester:** The name of the person(s) executed the test.
- **Test environment**: The location where the test took place.
- **Test date**: The date the test took place.
- **Actual result:** The results produced after the execution of the test.
- **Conclusion:** Whether the executed test was successful (pass) or not (fail).
- **Remarks:** Any important information derived during the test execution.

The use of test cases provides a clear testing process, ensuring thorough preparation, execution, and reporting. In addition, it helps to align all stakeholders and maintain focus on achieving quality objectives.

4.1.6 Testing Reviews

Testing reviews will be performed to ensure the quality of the units, through prompt, effective and efficient detection and addressing of problems. Their objective is:

- To check the testing output and verify its conformance to requirements
- To track and resolve possible non-conformities
- To keep the involved partners informed about the partial results, which should be taken into consideration during the next steps

Testing reviews will be organized by the WP7 leader and/or T7.3 leader and they will be performed after every integration and system tests with the participation of the partners that are directly connected to the units tested, as well as those that are dependent on or influenced by the specific units. The project coordinator will always participate to facilitate and guide the review. The results of each review will be recorded in a predefined format, will be distributed and kept for further use.

4.2 Test Plan

For the Testing Plan creation, several data required to be gathered by the partners. In this framework the key components described below have been identified. An initial version of the template to be distributed can be found in Annex I. Upon data collection, an analysis will be performed to produce a testing plan per testing site, lab environment and pilot sites.

- **Test number:** A unique number for the test to be able to easily track each test.
- **Test description:** A small description about what the test concerns.
- **Test items:** List with the units, modules, prototype or system and the features that will be tested.
- **Test objective:** The test goal. In accordance with the test framework, four options exist; "Functionality," "Performance," "Reliability", "User acceptance" and "Safety."
- **Test stage:** The testing stage. In accordance with the test framework, four stages exist; "Unit," "Integration," "System," "Validation."

- **Testing site**: The location where the test will take place, i.e. lab environment or pilot site.
- **Set up:** The set up that will be used for the test, i.e. airport or supermarket.
- **Use case:** The UC(s) the test relates to.
- **Partners involved:** List with the partners that are required to perform the test.
- **WP related:** The WP the test relates to.
- **Restrictions:** Mention any restriction the test has, e.g. can only be performed upon final testing of another module.
- **Testing duration:** The estimated duration required to perform the test.
- **Estimated test start date:** The desired date that the test should start.
- **Estimated test end date:** The desired date the test will have been completed.

5 Description of validation criteria and KPIs

Key Performance Indicators (KPIs) are essential metrics used to evaluate the success and efficiency of the MANiBOT project. They provide a quantitative basis for assessing both the technical and operational aspects of the robot's development and deployment. The KPIs used for MANiBOT project will be divided into two main categories:

- **Use Cases KPIs :** These KPIs are designed to measure the robot's performance in each specific use case, in retail and airport environments, ensuring that the MANiBOT meets the required operational standards. Such KPIs are described in D2.2 and can be further enriched during the project.
- **Technical KPIs :** These KPIs focus on the technical specifications and performance of individual robot components. They include metrics such as accuracy, reliability, efficiency and safety, ensuring that each component meets the standards required for the successful integration into the overall system. These KPIs are described in D2.4 and can be further enriched during the project.

In the updated version of the deliverable, namely D7.5, having gained deeper insights of the MANiBOT system and after clarifying the detailed use case scenarios to be implemented at the pilot sites, the complete and final list of the KPIs will be provided.

6 Risk & Issue Management Plan

6.1 Risk Management Plan

The MANiBOT Risk Management methodology is described in D1.1, paragraph 3.13. For WP7 works, risk analysis will involve a qualitative analysis in order to obtain the risk matrix based on likelihood and severity [3].

6.1.1 Definitions

Likelihood is the probability of occurrence of an event which is evaluated based on history (similar projects) and conditions for occurrence like context or measures adopted. The below Table presents the likelihood level scores.

Table 4 Likelihood level scores

Likelihood	Score	Description
High (H)	3	More likely to happen than not
Medium (M)	2	Fairly likely to happen
Low (L)	1	Unlikely, but not impossible to happen

Severity is the impact of an event on the project which is evaluated regarding the influence on project cost and schedule. The below Table presents the severity level scores.

Table 5 Severity level scores

Severity	Score	Description
High (H)	3	If the risk event occurs, the project will encounter major cost/schedule increases. Minimum acceptable requirements may not be met. Most secondary requirements may not be met. (This risk event may cause the project to fail.)
Medium (M)	2	If the risk event occurs, the project will encounter moderate cost/schedule increases. Minimum acceptable requirements will be met. Some secondary requirements may not be met.
Low (L)	1	If the risk event occurs, the project will encounter small cost/schedule increases. Minimum acceptable requirements will be met. Most secondary requirements will be met. (This risk event also may have no effect on the project and all requirements will be met.)

Risk level is the likelihood that a particular source threat exploits a vulnerability, as well as the impact of that event on the project which is evaluated using the following formula:

Risk level = Likelihood x Severity.

The risk score matrix based on the assessment of severity and likelihood is depicted in the Table below.

Table 6 Risk level scores

Risk Level		Severity		
		Low (L)	Medium (M)	High (H)
Likelihood	Low (L)	1	2	3
	Medium (M)	2	4	6
	High (H)	3	6	9

Depending on the risk level of each risk, the following actions will take place.

- Green/Negligible - Risks classified as green do not need additional mitigation actions and will be routinely tracked.
- Yellow/Important - Risks classified as yellow require mitigation actions. For these risks, alternatives will be identified and "tradeoffs" will determine the mitigation requirements.
- Red/Critical - Risks classified as red require very close attention regarding evaluation, mitigation actions and monitoring. For these risks, the mitigation actions will be developed. Red risks will be evaluated to determine the impact on budget and will be tracked down until they are closed, or they are under the acceptable risk level.

6.1.2 Risk treatment

The following approaches will be selected to address the WP7 risks that will be identified:

- **Avoidance/Prevention**: Terminate the risk - by doing things differently and thus removing the risk, where it is feasible to do so. Countermeasures are put in place that either stop the threat or problem from occurring or prevent it from having any impact on the project.
- **Reduction/Mitigation**: Treat the risk - take actions to control it in some way that the actions either reduce the likelihood of the risk developing or limit the impact on the project to acceptable levels. The mitigation actions will contribute to reduce the probability of occurrence. If mitigation actions do not provide expected results, then new mitigation actions will be determined.
- **Acceptance**: Tolerate the risk – when either if nothing can be done at a reasonable cost to mitigate it or the likelihood and severity of the risk occurring are at a low level.

6.1.3 Risk monitoring and review

A WP7 risk log will be kept by the WP7 leader to facilitate the risks monitoring, tracking and review. The relevant template can be found in Annex II. The WP7 risk log will be updated every month and will be presented during the WP7 monthly meeting. For each risk, a responsible, “risk owner,” will be assigned who will track, monitor, control and report on the status and effectiveness of each risk response action to the WP leader. The WP7 leader in coordination with the risk owner will take decision about risk closing, mitigation or continual monitoring.

6.1.4 WP7 risks

The WP7 risks will be identified at a later stage, and they will be included in D7.5.

6.2 Issue Management Plan

Apart from the risks, which are defined as potential situations that could occur in the future and plans and strategies can be created for their treatment, also project issues should be taken into account. The difference of issues from risks is that issues are defined as challenging events that do not refer to the future but have already happened or are happening requesting immediate action. These issues must be addressed as soon as they occur and for this during the course of the project, they will be recorded and closely monitored.

A WP7 issue log will be kept by the WP7 leader to facilitate the issues monitoring, tracking and review. The relevant template can be found in Annex III. The WP7 issue log will be updated every month and will be presented during the WP monthly meeting. For each issue, an "issue owner" will be assigned who will be responsible to perform the actions for the issue resolution, as well as report on its status to the WP7 leader.

6.2.1 WP7 issues

The WP7 issues will be identified at a later stage, and they will be included in D7.5.

7 Conclusions

In this phase of the project, the primary focus is the setting of the testing plan framework. As the deliverable is a high-level description, there are no concrete results or outcomes to present at this time. A future version will provide a more detailed analysis and insight as the project progresses.

Future actions include the definition of the test cases and the development of detailed test plans per testing site, lab environment and pilot sites, as well as the identification of WP7 risks and issues which will all be included in D7.5. In addition, after clarifying the detailed use case scenarios to be implemented at the pilot sites and determining the final KPIs, the MANiBOT evaluation procedure will be defined.

References

[1] NVIDIA, [Online]. Available: https://docs.omniverse.nvidia.com/isaacsim/latest/index.html. [Accessed 17 10 2024].

[2] [Online]. Available: https://gazebosim.org/home. [Accessed 27 10 2024].

[3] 29 10 2024. [Online]. Available: https://www.infodiagram.com/slides/3x3-risk-assessment-matrix-/.

Annex I: Template to gather data for creating the testing plan

Test No	Test description	Test items	Test Objective	Test stage	Testing site	Set up	Use case	Partners involved	WP related	Restrictions	Testing duration	Estimated test start date	Estimated test end date
			Functionality	Unit	Lab environment	Airport							
			Performance	Integration	Pilot site	Supermarket							
			Reliability	System									
			Safety	Acceptance									

Annex II: Risk log template

Risk ID	Risk Description	Risk Category	Risk Impact Description (Time / Cost / Quality)	Pre-Mitigation			Preventive – Mitigating Actions	Post-Mitigation			WP	Risk Owner
				Likelihood	Severity	Risk Level		Likelihood	Severity	Risk Level		
						0				0		
						0				0		
						0				0		
						0				0		
						0				0		
						0				0		
						0				0		
						0				0		
						0				0		
						0				0		
						0				0		
						0				0		
						0				0		
						0				0		
						0				0		
						0				0		
						0				0		
						0				0		
						0				0		
						0				0		
						0				0		
						0				0		
						0				0		
						0				0		
						0				0		
						0				0		
						0				0		
						0				0		
						0				0		
						0				0		
						0				0		

Annex III: Issue log template

Issue ID	Issue Description	1st Registration Date	Issue Impact Description	Issue Significance *	Actions	Issue Owner	Issue Status **	Issue Resolution Description	Planned Completion Date	Actual Completion Date	Comments
* Issue Significance: High: Project goals and objectives are highly impacted. Medium: Project goals and objectives are moderately impacted. Low: Project goals and objectives are poorly impacted or not impacted at all.						** Issue Status: New: No action has been taken yet. Investigating: The issue and possible solutions are being investigated. Running: Resolution is in process. Escalated: Escalation took place and directions or approval of a solution is pending. Closed: The resolution has been implemented and the issue is closed.					