

"THE HYPER EXPERT COLLABORATIVE AI ASSISTANT"

Deliverable 5.3

Report for Integration and Technical Validation (1st Version)

Version	2.0
Lead partner	FUJITSU
Dissemination level	Public
Type of deliverable	Demonstrator
Contractual date	31/06/2025
Actual delivery date	31/06/2025

History of changes:

Version	Date	Changes	Pages concerned
v1.0	Submitted on 31.03.2025	First draft	n/a
v2.0	Submitted on 30.06.2025	Included technical implementation and validation of MVP for DAT use case	p31-p42

D5.3. Integration and Technical Validation (1st Version)

This deliverable provides a 1st version of the demonstrator (MVP) of the conversation of human and AI showing a basic set of functions illustrating the principles of the conversation flow. Use cases will provide a partial data and messages integration for a limited communication flow, which demonstrates the abilities of the technology and facilitates further refinement in versions 2 and 3.

Authors	**Organisations**
Bertrand, Jan	FUJ
Neises, Jürgen	FUJ

Contributors	
Ehlert, Maike	AMS
Krishnan, Vishruth	AMS
Mendes, Matilde	SON
Borges, Frederico Guilherme	SON
Silva Moreira, Fabio	InT
Lunet, Miguel	InT
Princy, Andriampaniry	PDT
Floris Bleuland van Oordt	DAT
Joep Beurskens	DAT

Revised by	
Baier, Hendrik	TU/e
Nalepa, Gregorz	JU

Disclaimer
This document reflects the authors' views only and neither the Agency nor the Commission are responsible for any use that may be made of the information contained

Table of Contents

1.	**Executive summary**	**5**
1.1	Introduction	5
1.2	The PEER Project and its Objectives	5
1.2.1	Purpose of D5.3 Deliverable Report	5
1.2.2	Relation to WP5	5
1.3	Relation to WP3	6
1.4	Overview of the Use-Cases	6
1.4.1	Specific Goals of Each Use-Case	6
2.	**Description of Work**	**8**
2.1	Sonae	8
2.1.1	Motivation	8
2.1.2	Goals	8
2.1.3	Limitations	8
2.1.4	User Journey	9
2.1.5	Evaluation Journey	11
2.1.6	Design of the MVP	12
2.1.7	Responsibilities	15
2.2	City of Amsterdam	16
2.2.1	Goals	16
2.2.2	Limitations	16
2.2.3	User Journey	16
2.2.4	Evaluation Journey	23
2.2.5	Design of the MVP	23
2.2.6	Responsibilities	24
2.3	Proditec	24
2.3.1	Goals	24
2.3.2	Limitations	25
2.3.3	User Journey	25
2.3.4	Evaluation Journey	26
2.3.5	Design of the MVP	27
2.3.6	Responsibilities	30
2.4	Datacation	31
2.4.1	Motivation	31

2.4.2 Goals 31
2.4.3 Limitations 31
2.4.4 User Journey 32
2.4.5 User-AI Interactions 38
2.4.6 Socio Technical Requirements 39
2.4.7 Evaluation Journey 39
2.4.8 Design of the MVP 40
2.4.9 Responsibilities 42
3. Conclusions **43**

1. Executive summary

1.1 Introduction

The PEER project is a research and innovation initiative funded by the European Union's Horizon Europe program. The project aims to develop human-centric AI systems for sequential decision-making, focusing on enhancing user trust and acceptance through improved human-AI collaboration. This deliverable report, D5.3, focuses on the technical integration and validation of the PEER collaborative AI assistant in four distinct use-cases. It details the development and implementation of Minimum Viable Products (MVPs) for each use-case, highlighting the challenges and successes in integrating the AI assistant into real-world scenarios.

1.2 The PEER Project and its Objectives

The PEER project is driven by the compelling goal of addressing the lack of user trust in AI systems, a significant barrier to the widespread adoption of AI technologies. The project aims to achieve this by:
Improving User Understanding of AI: Enhancing transparency and explainability of AI decision-making processes to foster user trust.
Improving AI Understanding of Users: Tailoring AI solutions to user needs and preferences through mixed human-AI initiatives.

1.2.1 Purpose of D5.3 Deliverable Report

The D5.3 deliverable report specifically addresses the technical implementation and validation of the PEER collaborative AI assistant within four distinct use-cases. The report details the development and integration of Minimum Viable Products (MVPs) for each use-case, highlighting the progress made towards achieving the project's objectives.
The purpose of this report is to:

- Document the technical specifications and design of each MVP.
- Demonstrate the integration of the AI assistant into existing systems and workflows.
- Present the results of the technical validation process.
- Identify challenges encountered and lessons learned during the MVP development.

1.2.2 Relation to WP5

Work Package 5 (WP5) of the PEER project is dedicated to "Learning: Piloting and Evaluation." WP5 focuses on implementing, applying, and validating the PEER collaborative AI assistant in four use-cases across different sectors. The development and evaluation of MVPs, as detailed in this D5.3 report, are core activities within WP5, contributing to the overall goal of integrating and validating the AI assistant in real-world scenarios.

1.3 Relation to WP3

The general architecture of the use cases follows the building blocks of the PEER Companion Agent as described in D3.1 and illustrated by Figure 1 below.

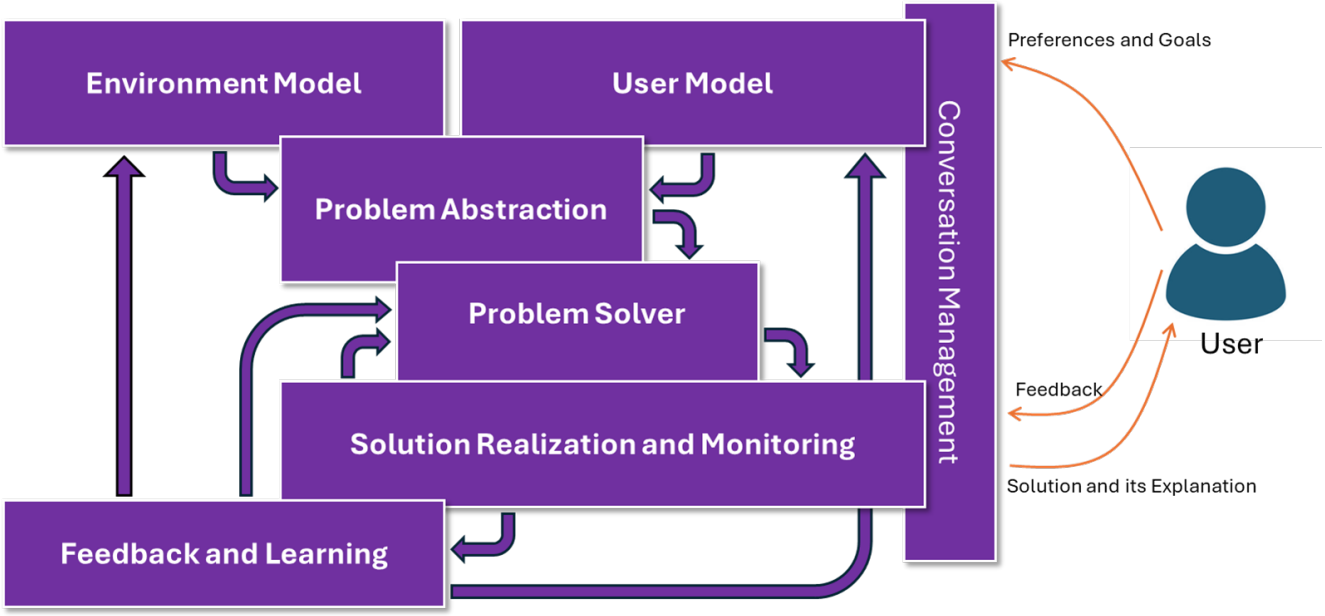


Figure 1 Architecture of the PEER Companion Agent.

The use case specific architectures are depicted in Description of Work For more details of use case specific architecture please, refer also to D3.1.

1.4 Overview of the Use-Cases

The PEER project has identified four use-cases to demonstrate the capabilities of the collaborative AI assistant:

- Retail: Personalized in-store picking assistant for customers.
- Smart & Inclusive City: Personalized routing systems for people with disabilities in Amsterdam.
- Quality Control: Image-based quality control of manufactured products. (Coins or Pharmaceutical products)
- Logistics Optimization (Datacation): AI-assisted optimization of warehouse location type allocation based on future orders.

1.4.1 Specific Goals of Each Use-Case

- Use-Case 1: Retail - SONAE
  Develop an AI-based personalized in-store picking assistant for customers.
  Optimize picking routes based on product location, weight, category, and customer preferences.
  Enhance customer experience and satisfaction.

- Use-Case 2: Smart & Inclusive City - City of AMSTERDAM
  Develop an AI-based personalized routing system for people with disabilities in Amsterdam.
  Consider various accessibility criteria, such as sidewalk width, step height, and obstacles.
  Improve the mobility and autonomy of people with disabilities.

- Use-Case 3: Quality Control - PRODITEC
  Develop an AI system that can identify and classify manufacturing defects in coins.
  Reduce the manual effort required for quality control.
  Improve the efficiency of the quality control process.

- Use-Case 4: Logistics Optimization – DATACATION
  Develop an advanced AI system to assist logistics engineers in optimizing warehouse location type allocation. Prioritize collaborative Human-AI interaction for decision-making based on future incoming orders.

2. Description of Work

2.1 Sonae

2.1.1 Motivation

It is hard to develop a customer choice model that accurately mimics customer behaviour within the store environment. It would require understanding:

- customer preference for all products
- customer willingness to walk more or spend more time in a store
- customer willingness to wait in line at the fish and meat stations
- customer willingness to leave the store.

Due to this complexity, the best way to grasp customer dynamics is through a workshop in which the customer will interact with a simulator (conclusion from Porto meeting in September).
The simulator has been integrated into a web app. In further steps in can be extended towards a mobile phone app, to closer mimic the already existing SONAE customer app, Continente Siga. The higher the proximity to the real-life scenario, the better the results will be.
The simulator should be a tool that allows testing the developed policies regarding routing (MVP1) and promotion suggestions based on customer preferences (next MVPs).

2.1.2 Goals

The simulator overall goals are:

1. **Virtually represent the store environment**
    - Products must be assigned to a location
    - Product characteristics should be assigned to the product (price, weight, frozen/fresh, category, etc.)
    - The waiting time for meat and fish station queues must be tracked and can be a source of uncertainty
2. **Mimic customer dynamics**
    - Customers can update their shopping lists and baskets
    - Customers can impose restrictions (precedence constraints for product purchase, maximum shopping time)
    - Customers are able to purchase products
    - Allow for picking policy testing
3. **Store customer behaviour data**

2.1.3 Limitations

The main limitation of MVP 1 is that it does not provide the experience we want to offer users of the already functioning Continente Siga App. To effectively test with real users, we need a **front-end** that mimics the Siga App experience. Given this constraint, for this MVP, we will conduct **testing**

exclusively with internal Sonae users. However, in future MVP iterations, **Fujitsu considers development of a dedicated front-end** that better replicates the Siga App experience using wireframes already designed by Sonae. This will bring the testing environment closer to real-world conditions and, hopefully, enable us to test with actual users.

In addition to this limitation, MVP 1 does not feature **real product locations**. It represents main category aisles in a simulated environment. To achieve product locations in a real store in future MVPs, we are collaborating with the **INESC TEC robotics team**. Furthermore, in the current version, users must **manually navigate** within the simulator. In future MVPs, the system should automatically detect user movement based on indoor location and routing and adjust the route, accordingly, providing a more seamless and realistic experience. However, a solution for indoor location and routing has not yet been specified.

2.1.4 User Journey

As shown in Figure 2, the user journey can be outlined in the following steps:

1) The user enters the simulator and selects his preferences by:
   - **Adding products to his/her shopping list**
   - **Selecting the shopping target** (minimize shopping distance or minimize shopping time) and **setting precedence constraints** (e.g. Picking the frozen products last)

2) Once the user submits the preferences, **a route is generated**. The user can change the picking order and add products to his list at any time during the shopping journey;

3) The user can then **simulate navigation** through the store. For that the user can follow or deviate from the proposed route (leading to a new route being generated) as well as **update the shopping list** and **pick products;**

4) Once the user **reaches the checkout**, his simulated shopping journey ends and is presented with a review of the shopping journey;

5) The user can then **rate the experience** from 1-5 in the simulator and provide qualitative feedback.

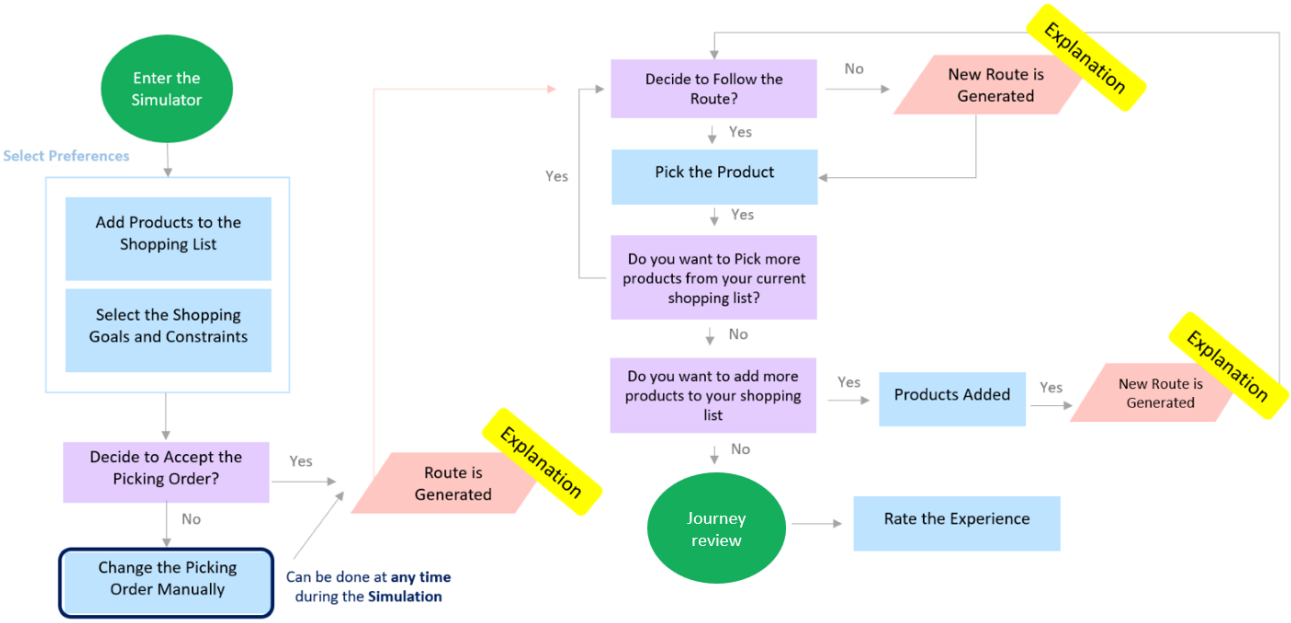


Figure 2 SONAE User Journey in MVP

2.1.4.1 User-AI Interactions

For this first MVP, as mentioned above, the user initially interacts with the system by selecting the shopping list and preferences. Once a **route is generated**, a **pre-defined explanation** is given to the user (e.g., "This route minimizes your shopping time to a total of 45 minutes"). The user can then follow the proposed route or deviate. If he deviates, and a **new route is generated**, and the user receives **another pre-defined explanation** (e.g., "By deviating, your shopping time has increased 8 minutes"). At the **end of the journey**, the user is given a review of the journey (e.g. Total Shopping Time) and **interacts again** with the system by **rating his/her experience.**
With the MVP testing we hope to understand what other **explanations** would be useful to add value to the shopping experience and enhance the trust and transparency in the system.

2.1.4.2 Socio Technical Requirements

Table 1 SONAE Socio Technical Requirements

Question	Answer
What parts/features of the system do you want to test?	Test a simulator that supports the customer through the entire shopping journey including the creation of the shopping list, the initial conversation/interaction and suggestion of route, navigation, and post-shopping feedback.

Why do you want/plan to test it? What do you hope to learn (testing goal)?	• We want to learn customer needs regarding the shopping journey (e.g., does it make sense to have a route-based aid tool?). • We want to understand what type of feedback can be or should be collected during a shopping journey without disturbing the shopping experience.
With whom do you want to do the test (e.g., end-users)?	Internal users in a controlled environment (e.g., a workshop).
When do you want/plan to do the test (timing)?	By the end of April 2025.
What tools do you have available to perform the test (e.g., Figma, wireframes, nothing)?	We will have a simulation environment that enables this type of interaction with the retail customer. We also have some wireframes that can be used if needed as a complement to the simulator.
What will be (technically) working in the MVP (e.g., AI module)?	We will have a clickable simulation environment, like a game, where the customers can follow a proposed shopping path that is based on their shopping list, selected shopping objective and precedence constraints. These shopping paths will be proposed by operations research algorithms that do not consider user feedback, thus, in terms of AI, we believe nothing will be available by the first MVP.

2.1.5 Evaluation Journey

The evaluation phase will focus on **gathering insights from user interactions** within the simulated environment. Key evaluation points include:

- User adherence to the suggested routes and reasons for deviations
- Time spent on shopping and the impact of constraints (e.g., preference for picking frozen products last)
- Flexibility of the routing algorithm for different shopping objectives
- User feedback on route clarity, convenience and how that leads to trustworthiness
- Potential friction points and positive points in the customer experience

An example of an evaluation screen available in the current MVP simulator is shown in Figure 3.

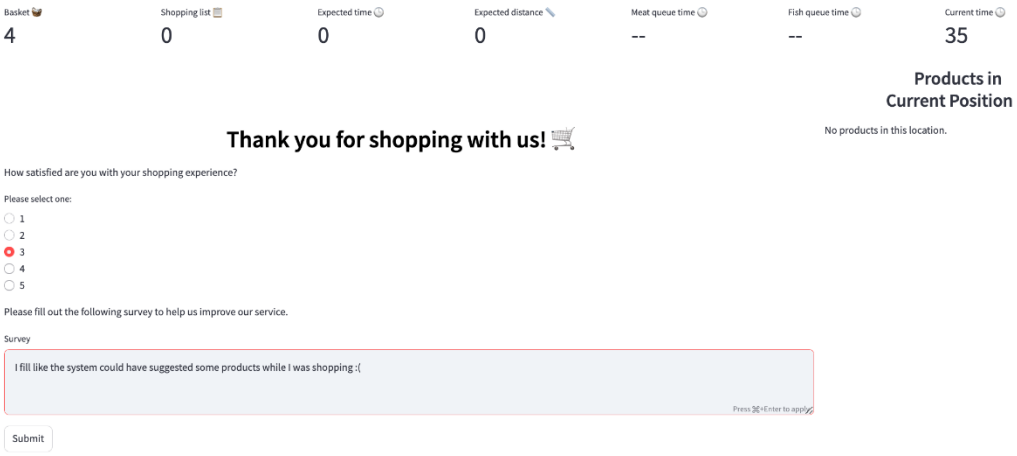


Figure 3 SONAE Evaluation Screen

2.1.6 Design of the MVP

The system includes a set of modules, presented in Figure 4, which are described in the following subsections.

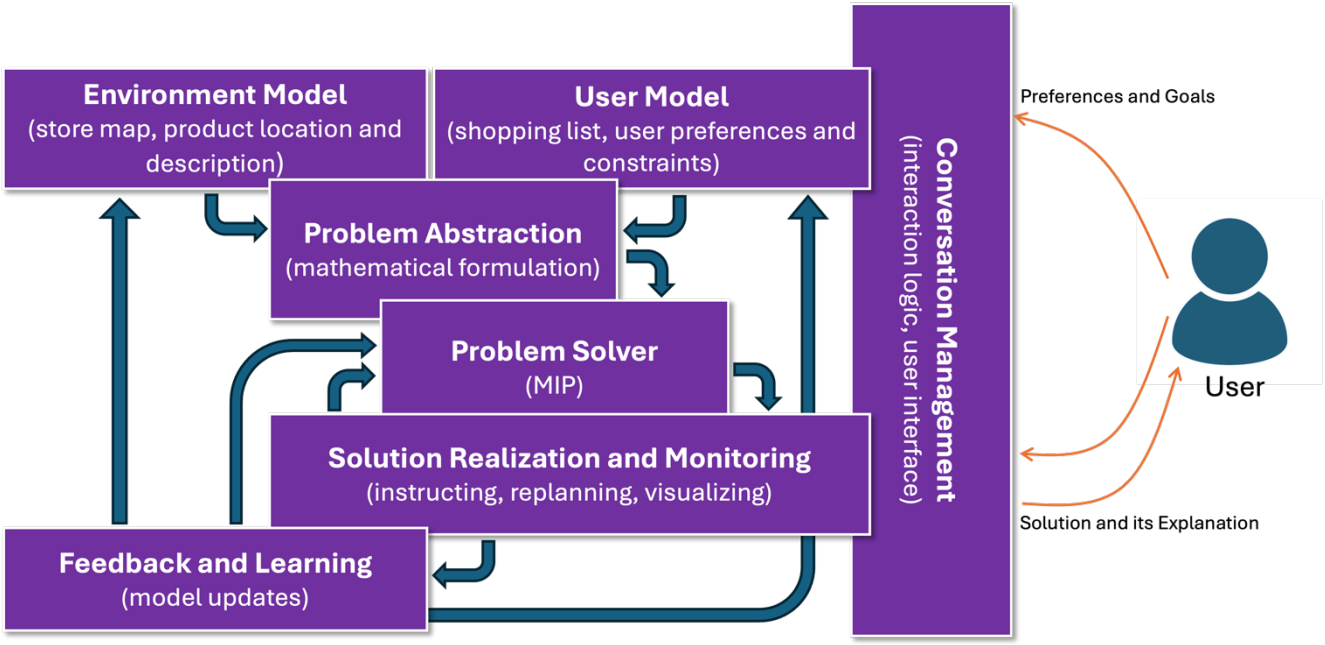


Figure 4 PEER building blocks mapping to personalized in-store shopping assistant use case

2.1.6.1 Design choice due to technology challenges

Given that the real-world integration of SIGA's new routing system is not yet feasible, we opted for a simulation-based testing approach. The following constraints influenced our design:

- Lack of real-time in-store user tracking
- Limited customer feedback data for preference learning
- Computational complexity of real-time route optimization

2.1.6.2 Product Locator

One of the biggest challenges in worldwide retail nowadays is to maintain a digitalization of the store with sufficient detail to describe the product locations in every moment. This is very challenging due to the constant changes occurring in the stores (e.g., promotional activity). INESC TEC is developing an approach that scans store aisles in order to generate a graph including the product locations inside the retail store. This is a key step to obtain the necessary data (Environment Model represented in *Figure 4*) to run route planning algorithms.

2.1.6.3 Preference Extractor

The initial customer interaction gathers preferences and constraints such as preferred shopping objectives (e.g. Fastest Route), time constraints, and specific product priorities (User Model represented in *Figure 4*). This information is intended be used to generate an initial route suggestion using advanced routing algorithms. In the first MVP, these preferences are directly introduced by the customer in the form of constraints. The data regarding these choices will be collected and used in the future PEER approaches. Furthermore, in the next MVPs, these constraint preferences will be

integrated with past shopping lists to further personalize picking routes suggested (Feedback and Learning represented in *Figure 4*).

2.1.6.4 Static Router

The static routing algorithm (Problem Solver represented in *Figure 4*) computes an optimal path for customers based on store layout, product locations, customer preferences, and constraints such as precedence requirements (Problem Abstraction represented in *Figure 4*). In this first MVP we are using a MIP-based approach, meaning that, although we consider several business constraints, we can only handle small shopping lists. Indeed, the mathematical formulation can be solved by the commercial solver (Gurobi) only for small shopping lists. This routing approach is also used, myopically, to re-schedule the picking sequence when the customer deviates from the initially proposed route. In the next MVPs we intend to explore solution approaches that are better for sequential decision-making (e.g., Direct Lookahead Policies (DLA), Reinforcement Learning (RL)).

One interesting feature we included in the first MVP is the possibility to provide simple explanations regarding the routes suggested. For example, when the user chooses to pick frozen products in the last position of the route, we are able to tell him how much time he is losing for adding that constraint.

2.1.6.5 Simulation Environment

The simulator is designed as a gamified interface where users can follow the suggested route or deviate from it, modify their shopping list mid-journey, experience real-time re-routing based on shopping list updates or new constraints, and provide post-shopping feedback on the route experience (Solution Realization and Monitoring + Conversation Management). This environment serves as a controlled testbed to refine the routing logic and prepare for integration into SIGA in future MVP iterations. Figure 5 shows an example of the current simulation environment, which was developed using the Streamlit open source Pyhton library



Figure 5 SONAE Simulation Environment

2.1.7 Responsibilities

During this first period, the research team mainly focused on defining the requirements for the retail use case and building the necessary modules to tackle it. A use case Gantt chart was proposed, including tasks to achieve a first version of most modules to figure in the final approach. The main objective was to build a first pipeline which could extract customer preferences, propose static picking routes based on customer-defined constraints, and simulate the execution of a picking route inside a retail store. In parallel, the research team is developing a methodology to adequately locate products inside retail stores. All the tasks are considered to be on track, since the ones that have short delays are not in the critical path of this phase.

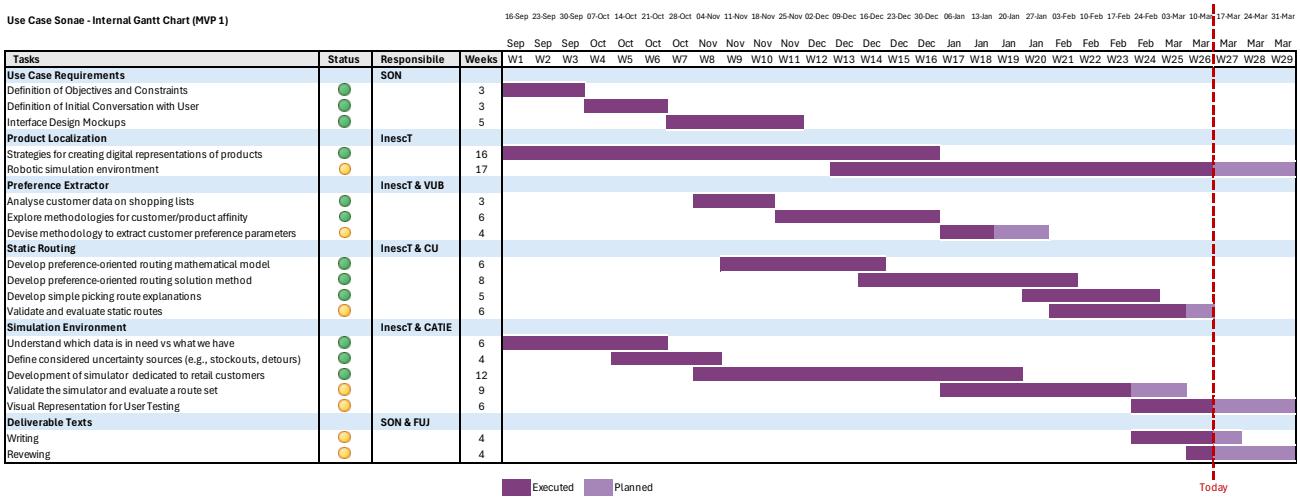


Figure 6 GANTT Chart with Partner Responsibilities and Timeline

Table 2 Partner Responsibilities and Status

Modules	Status	Explanation	Responsible
Use Case Requirements	Finished	The definition of the use case was successfully achieved in the first month of the project.	SONAE
Product Localization	In progress	The product location methods are substantially progressing and on track for this first period.	INESC TEC
Preference Extractor	In progress	The methodologies are being implemented and tested. There are significant delays in this task but it is not disturbing progress of other tasks.	INESC TEC & CU
Static Routing	In progress	Validation tasks are still open but on track.	INESC TEC & CU & VUB
Simulation Environment	In progress	Validation tasks are still open but on track.	INESC TEC & CATIE
Deliverable Texts	In progress	No delays are reported.	SONAE & FUJITSU

2.2 City of Amsterdam

2.2.1 Goals

- Develop a route planning system capable of generating and explaining diverse routes based on user preferences and constraints, with a focus on people with special mobility needs.
- The goal is to have precomputed routes, precomputed explanations, and wireframes depicting the routes by 31.03.2025.
- We also test different hypotheses for different explanations, visualizations of routes, etc.

2.2.2 Limitations

The testing is being done at the prototype phase. Thus, it's not market ready and not a typical MVP as understood in business context, it's an MVP for lab evaluation. Additionally, the routes are precomputed. The set of choices made by the user are already included (since the route needs to be pre-generated) so the user will be unable to make changes to the preferences, such as settings of origin and destination address. However, they could indicate preferences between a walk and bike path and other choices as detailed later.

We would also need to conduct a data privacy impact assessment and implement measures accordingly. The prototype has been built within a secure test environment using Figma and pre-computed routes based on open data. There is a chance we miss potential insights which could be gathered from users if the prototype was ready to be tested in the public space. But for this MVP we are not focussing on testing the prototype in the public space. These could be opportunities for the future. These insights could then relate to how a potential real-time interaction could work or other general feedback along the route.

2.2.3 User Journey

- The user is presented with a set of choices that impact the route offered. A more detailed explanation of the different steps is offered in the following section 2.2.3.1.
- The routes are pre-generated.
- The MVP focuses on route preparation at home, not on navigation during the route or learning from user behaviour during the route. Some post-navigation feedback options are tested as well. We focus only on testing the interactions.
- The MVP has the same starting point and destination for all users.
- The MVP will test user profile, daily preferences, and suggest alternative routes.

2.2.3.1 Presentation of the routes

The overall presentation of the routes follows the flow as shown in Figure 7. The Figma mock-up starts off with questions on general preferences of the user, followed by the presentation of a route and this is followed by possibility to make further adjustments. However, the goal of the MVP is to also present a few different forms of how the routes are presented. The general flow is discussed below. Starting with an onboarding process (which gathers information on the user – user profile), followed by

preference for speed, proximity to Public Transport Stops and lastly if there's a preference between a walk or a bike path.

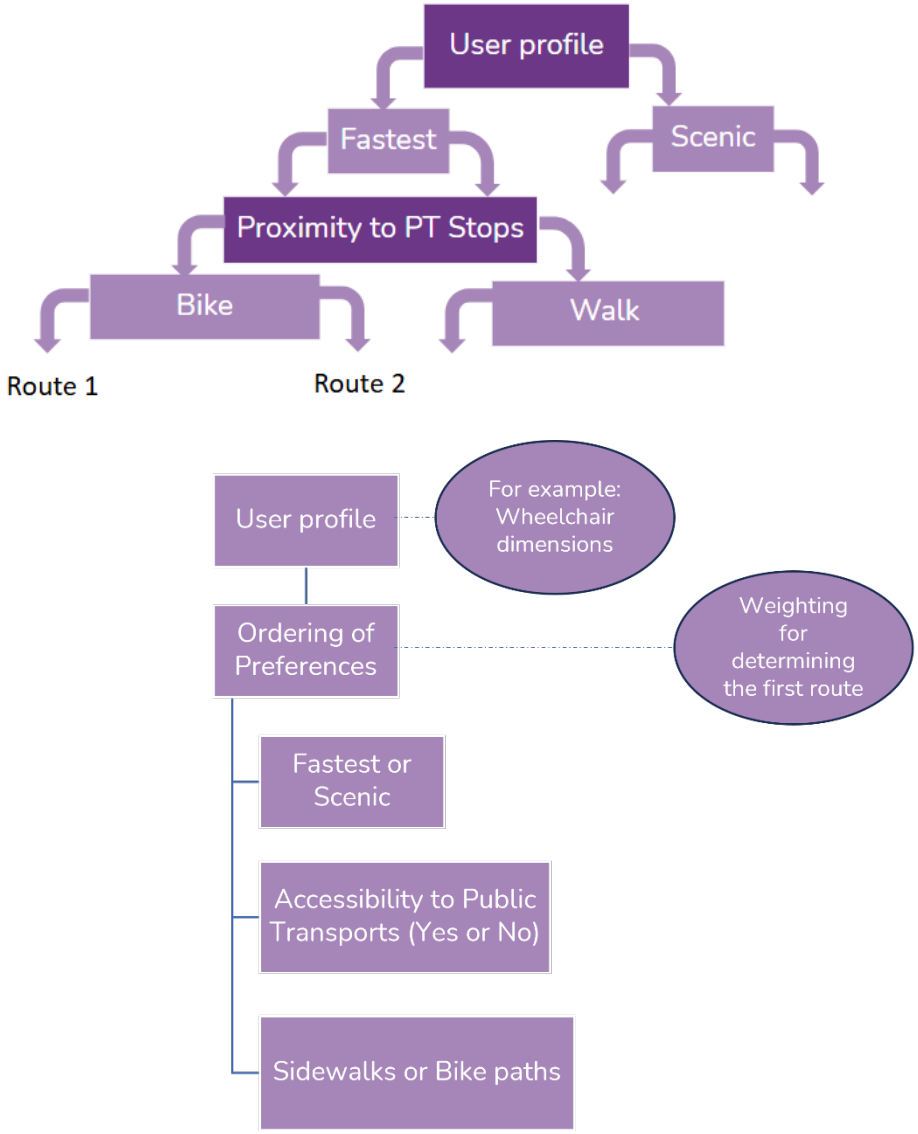


Figure 7 City of AMSTERDAM General flow of the MVP

The onboarding process involves asking general questions related to the wheelchair the user has, such as the type of wheelchair and its dimensions. These are shown in Figure 8 and Figure 9.

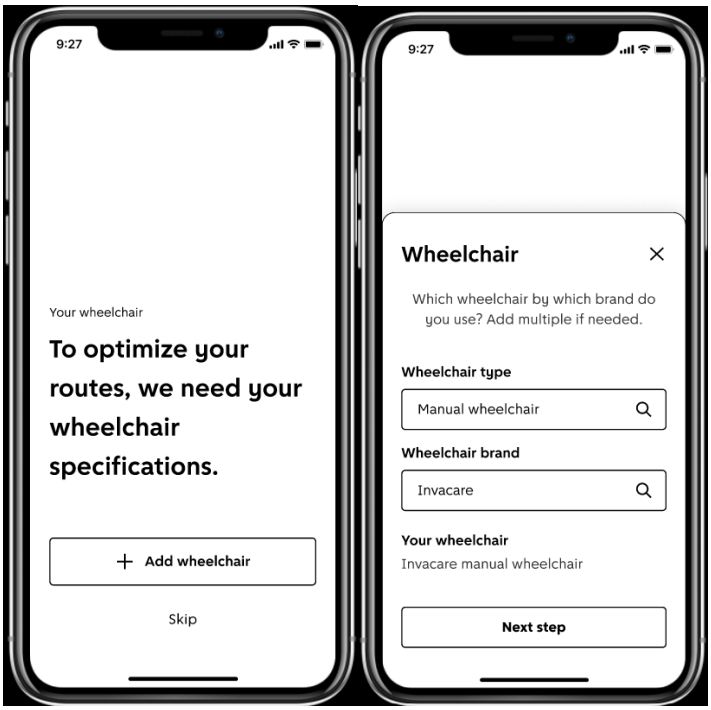


Figure 8 City of Amsterdam Wheelchair specifications - Onboarding

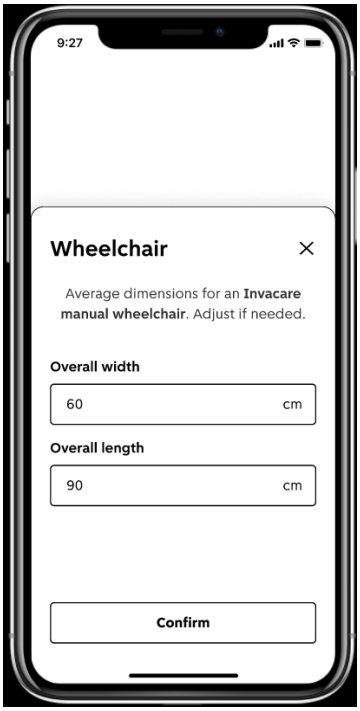


Figure 9 City of Amsterdam Wheelchair Dimensions- Onboarding

This is followed by the following questions on:

- Whether the route needs to be 'fast' (shortest possible time) or 'scenic' (across green spaces).
- Whether proximity to accessible public transport stops needs to be maintained (for example if the user needs to be always within a few hundred meters of a public transport stop).
- Lastly, if the user has a preference to use bike paths over footpaths.

The wireframes for gathering the information are shown in Figure 10.

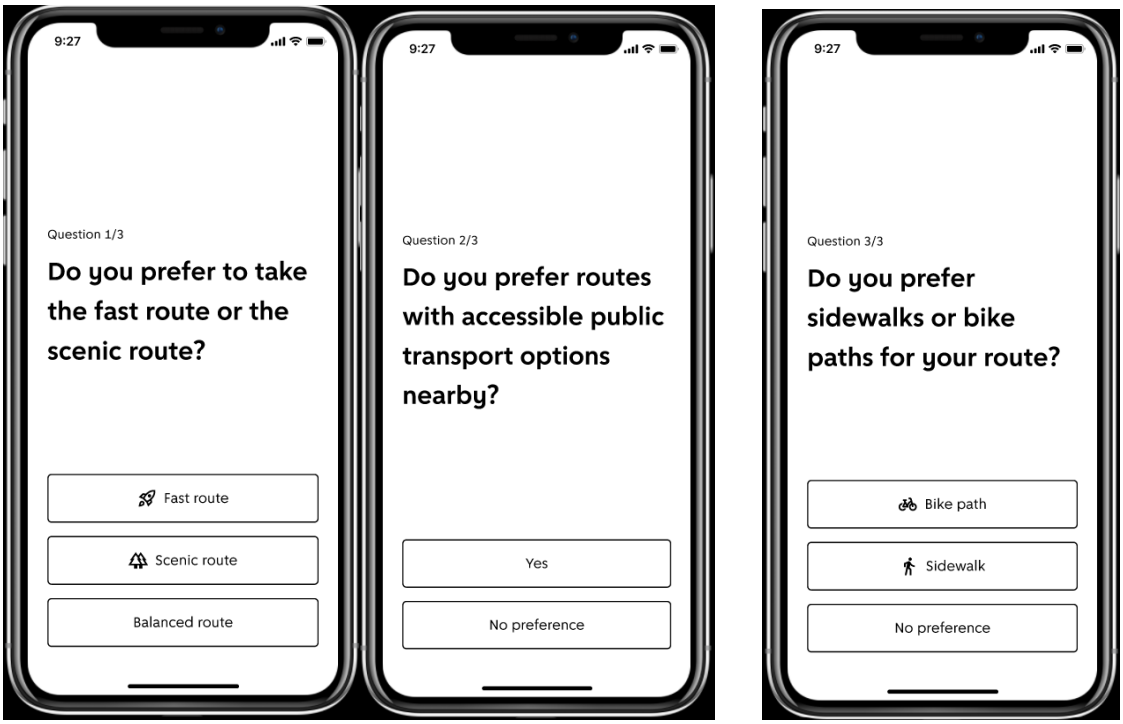


Figure 10 City of Amsterdam Questions around preferences

These preferences are already included in the baseline route planner to generate the routes. The algorithm has generated routes based on these user inputs and these have been displayed in the wireframes. Additionally, Figure 11 shows the order of priority for the routes which the user can change. This served as weighting for the algorithm.

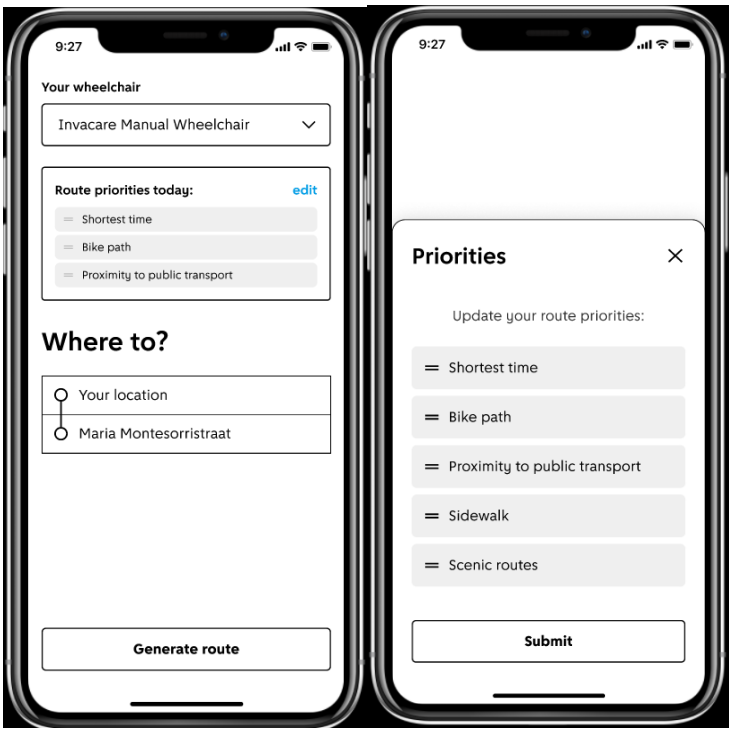


Figure 11 City of Amsterdam Specifying origin/destination and ranking preferences

This is followed by the visualization of the route, an initial route is shown followed by a screen showing a route with a reduced number of crossings, but the user must take a slightly longer route. These screens are visualized in Figure 12. We simulate the decision; the user has chosen for reduced number of crossings.

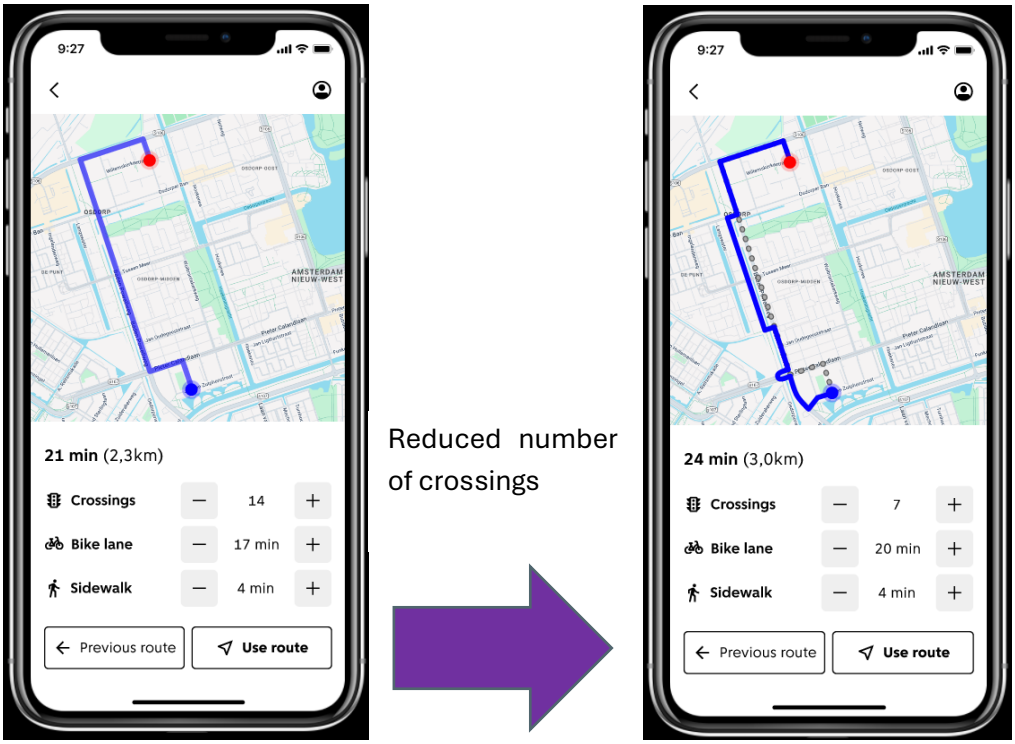


Figure 12 City of Amsterdam Routes generated followed by an alternative route with reduced number of crossings

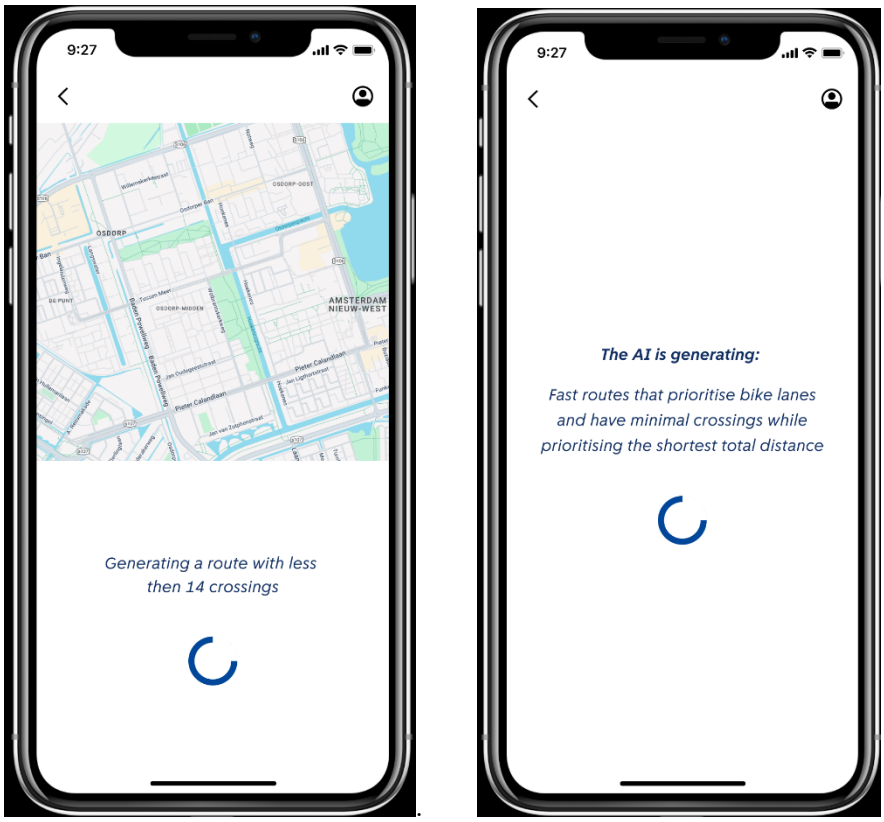


Figure 13 City of Amsterdam Loading screens with explanations

During the calculation of the routes, we show a loading screen explaining what is being calculated (Figure 13).

In addition to the screenshots included, there are also screens which offer the user a chance to give feedback on the route presented. This includes a rating and also edges or segments of the routes which should not be included or suggested in the future. The wireframes for how the routes can be rated are shown below in Figure 14.

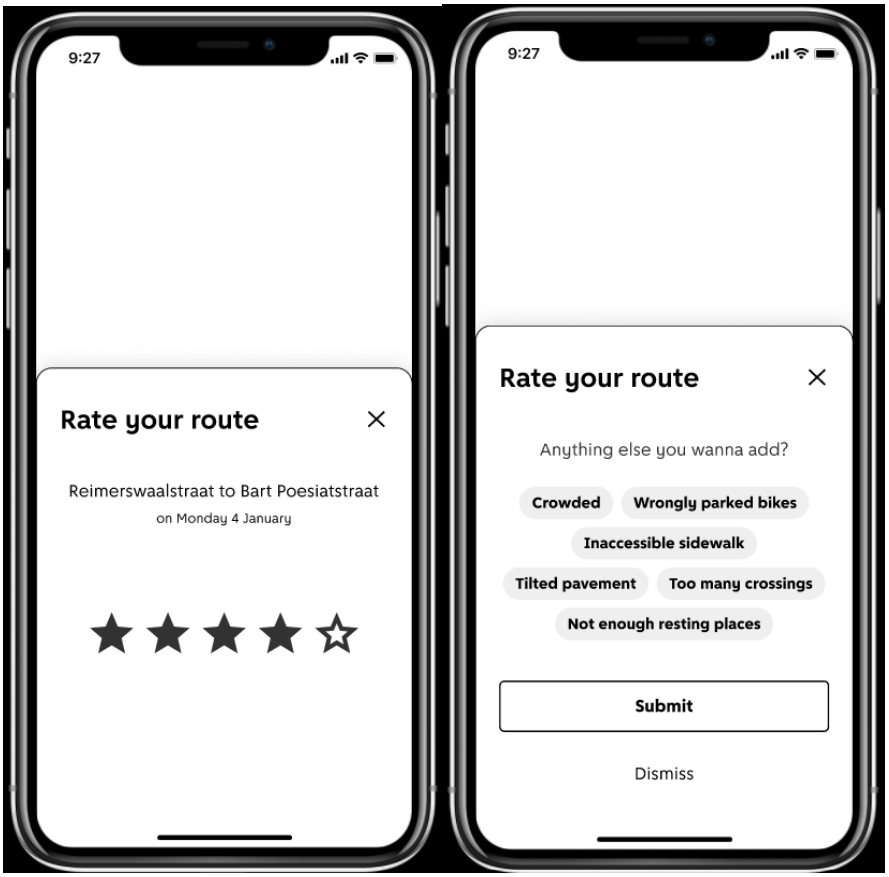


Figure 14 City of Amsterdam Example of how the feedback is asked in the MVP

2.2.3.2 Socio Technical Requirements

Table 3 City of Amsterdam Socio Technical Requirements

Question	Answer
What parts/features of the system do you want to test?	Testing if users are willing to take the time to offer more information on their preferences and their satisfactions with the newly presented alternatives. There's also a testing of how much information can be presented on the screen and how can it be presented. See all assumptions we suggested in Appendix 1.
Why do you want/plan to test it? What do you hope to learn (testing goal)?	• The MVP will gather potential new preferences and additional features. • The MVP will gather if users are interested in having more personalized routes. • The MVP will test how long users are willing to wait for route generation.

	• The MVP will test how much information is the user willing to digest to get better personalized routes.
With whom do you want to do the test (e.g., end-users)?	Wheelchair users, in collaboration with Clientenbelang Amsterdam
When do you want/plan to do the test (timing)?	29 & 30 April 2025
What tools do you have available to perform the test (e.g., Figma, wireframes, nothing)?	Figma, wireframes
What will be (technically) working in the MVP (e.g., AI module)?	The routes will be pre-generated based on an algorithm (sub-gradient), which generates routes based on different constraints. But as such no AI component will be working.

2.2.3.3 User-AI Interactions

For the first prototype we want to test with the end-user during which steps they would like to interact with the route planner. Are they willing to interact with the system at the start, during or/and at the end of the route.

2.2.4 Evaluation Journey

- Test user profile, daily preferences, route alternatives, final route usability, and feedback.
- List potential new preferences and additional features.
- Test user tolerance for waiting time during route generation.
- Conduct lab testing with a group of wheelchair users from Clientenbelang Amsterdam, using the local language and clickable wireframes.
- Test how many explanations and options are good enough for the end-users.

2.2.5 Design of the MVP

- The MVP will consist of modules for route generation, user profile, and human explanations.
- Route generation will use existing city data sources and precomputed maps.
- The datasets used in the MVP are based on the point cloud data, topographical data, Open Street Map data and Project Sidewalk data fused together to provide information in the form of a graph indicating location of crosswalks, accessible widths, and height of the footpaths.
- The routes generated are determined based on a variety of preferences already indicated through user-research. For example: preference to limit the number of crosswalks or maintaining proximity to public transport stops.
- The routes were based on the sub-gradient algorithm.
- The user will make a sequence of decisions based on the routes presented and their characteristics.
- Prototypes were created in Figma. Different possibilities were designed, to compare different visualizations of routes, different ways of giving feedback, and different ways of onboarding.

- Human explanations were based on the generated routes.

The context within the PEER building blocks is depicted in the following Figure 15 as illustrated also in D3.1.

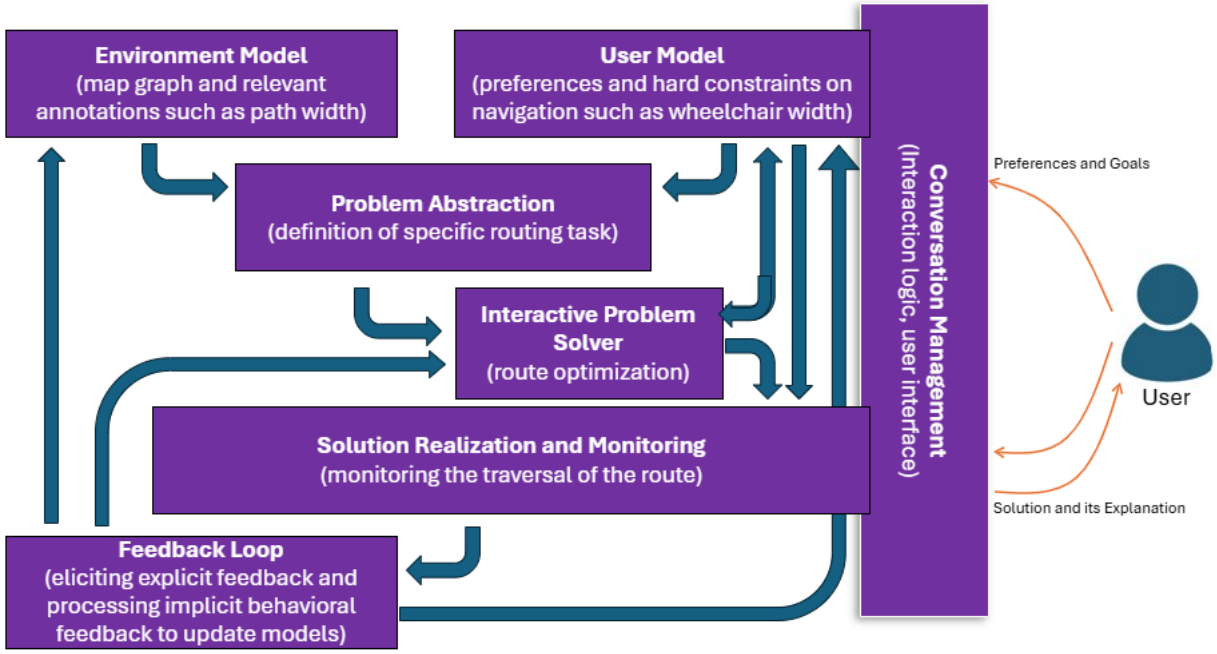


Figure 15 City of Amsterdam PEER building blocks

2.2.5.1 Design choice due to technology challenges

- We have a fixed starting point and destination. We do not have a live prototype for MVP1. We cannot compute new routes based on user choices. It is a design prototype.
- We created designs with the data we have available right now. However, the end-user may have further wishes to have into consideration on calculating the route.
- We designed an MVP that will be tested for this deliverable that takes into consideration the current technical possibilities (existing data and in development model). However, we created two more prototypes that share how different functionalities could look in the future to give a better impression on the potential of such personalised route planner.

2.2.6 Responsibilities

- AMS is responsible for route generation, and wireframe design.
- TU Eindhoven is responsible for the explanations.
- CATIE/SMIT-VUB and AMS are responsible for user testing.

2.3 Proditec

2.3.1 Goals

The objective of the Proditec MVP is to enhance the process of training anomaly detection models for image-based quality control for circulating coins.

The MVP aims to achieve this by:

- Improving transparency and user confidence in the AI system.
- Simplifying model training for users with varying skill levels.
- Increasing user understanding of the AI system's behaviour and defect detection capabilities.

The MVP will focus on the following:

- Developing a new software module, the **"AI model trainer."**
- Creating explanations and guidance for the model training process.
- Implementing AI or conventional transfer functions for model creation.
- Ensuring integrability with existing systems.
- Integrating and testing the new software module.
- Altering the front-end to accommodate new use-cases.
- Creating wireframes for the user interface.
- Evaluating the MVP using key performance indicators (KPIs).

2.3.2 Limitations

The current version of the prototype is built with mockups that replicate the behavior of the Peer assistant through interaction, but no calculations or API calls are currently being made on the assistant models, e.g. we simulate the entire process of auto-labelling without inferring the images through the model of peer assistant.
The goal of this MVP is only to test user interaction on the assistant modules without integrating the assistant into Proditec's final software. We will still need to evaluate the computational capabilities required to integrate the assistant into the hardware architecture of the machines. Other aspects such as traceability and user friendliness will be implemented and reworked during the integration of Peer assistant into Proditec's software (when the module is mature enough for the next iterations) because we do not yet have access to the databases and storage systems with the wireframes.

2.3.3 User Journey

The following outlines the typical user journey with the Proditec MVP and is depicted in Figure 16:

1. The operator enters the recipe setting for anomaly detection model training.
2. The operator reads explanations from the “AI trainer assistant” of all the steps to perform for training the model.
3. After image acquisition *(not simulated)*, the operator begins with the labelization task.

4. The semi-supervised module of assistant will help the operator to label the entire dataset, an adjustment from the user on potential error is only needed after the auto-labelling
5. The assistant asks the operator about his technical skills to adapt the HMI for the next steps
6. If the user is completely new to AI, we will use the "AutoML" module of the assistant to optimize hyperparameters, otherwise the software will give the possibility to configure them.
7. The operator initiates the training of the model.
8. After the training, the operator needs to configure a threshold depending on sorting requirements as the model produces anomaly score from images.
9. The explainability module of assistant will help the user to understand the defect detection through heatmaps.
10. If the accuracy is not enough at this stage, the operator needs to improve the dataset labelization using the active learning algorithm of the assistant.

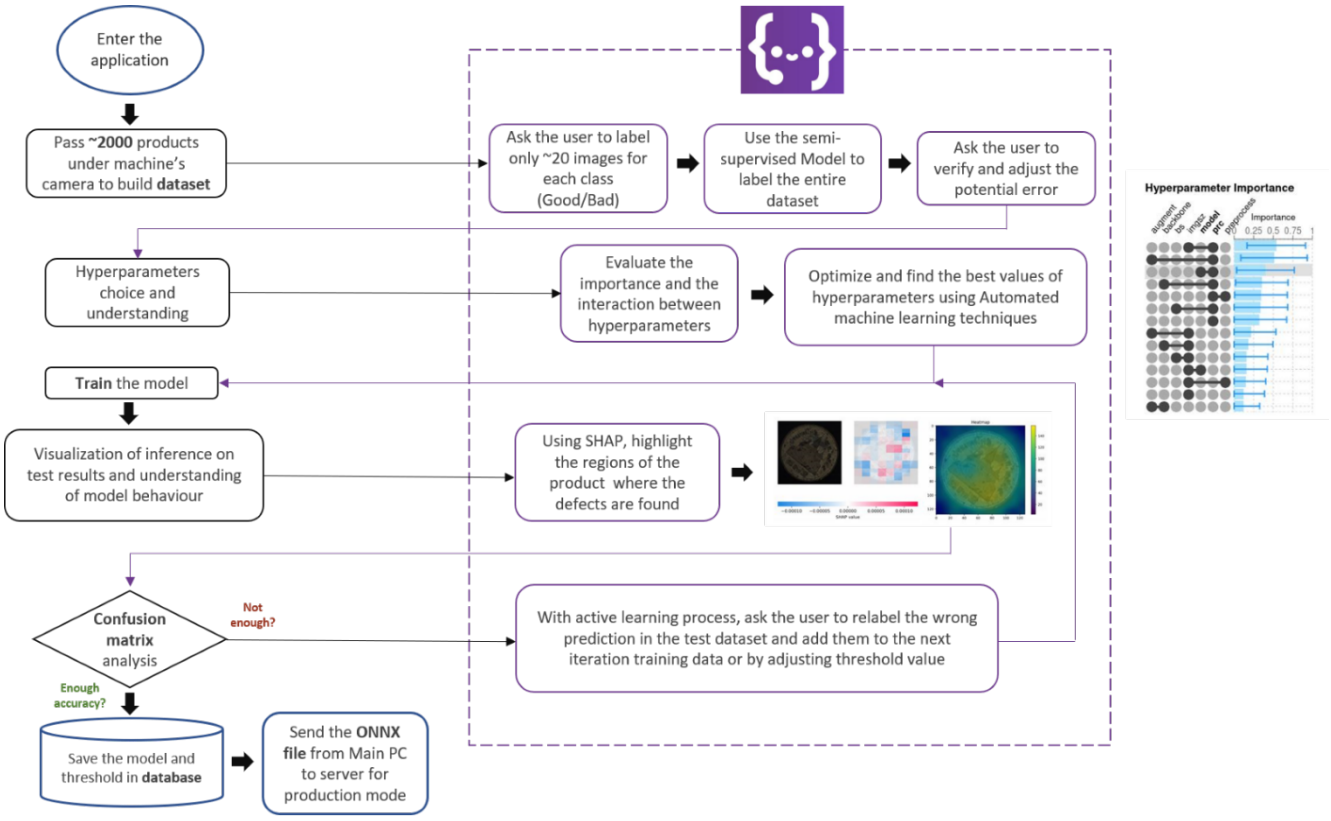


Figure 16 Proditec Process diagram

2.3.4 Evaluation Journey

The evaluation journey focuses on assessing the effectiveness of the AI model trainer companion in assisting operators with the model training process. The evaluation will involve:

- Comparing the operator's behaviour using the initial software (without the AI companion) versus a mock-up adapted with the AI companion functionalities, simulated by Wizard of Oz (WoO).

- Testing the effectiveness of the AI companion as a whole and analysing the operator's behaviour during each step using event logs.
- Evaluating how the explanations, suggestions, and clarity of the process help the user make decisions to achieve the accuracy goal.

2.3.5 Design of the MVP

The Proditec MVP consists of five technical key modules:

- **Guidance**: Provides guidance and explanations throughout the model training process, using mainstream terms for all user profiles.
- **Parameters optimization**: Automated Machine learning module for hyperparameters setting.
- **Semi-supervised Model**: Implements a semi-supervised model for image classification to pre-label images, reducing manual labelling effort. This module will be researched in the upcoming iteration.
- **Active Learning**: Integrates active learning to help the AI model converge to an optimal solution, allowing the AI companion to query the user for annotating new data or for the correction of ambiguous predictions.
- **Explainability Layer**: Improves the explainability of defect detection through heatmaps, increasing the understanding of the anomaly detection model's behaviour.

These modules are aligned to D3.1 in Figure 17. However, the specific technical modules overlap and combine tasks of the building blocks. Guidance, for instance includes elements of the User Model as well as of the Feedback Loop:

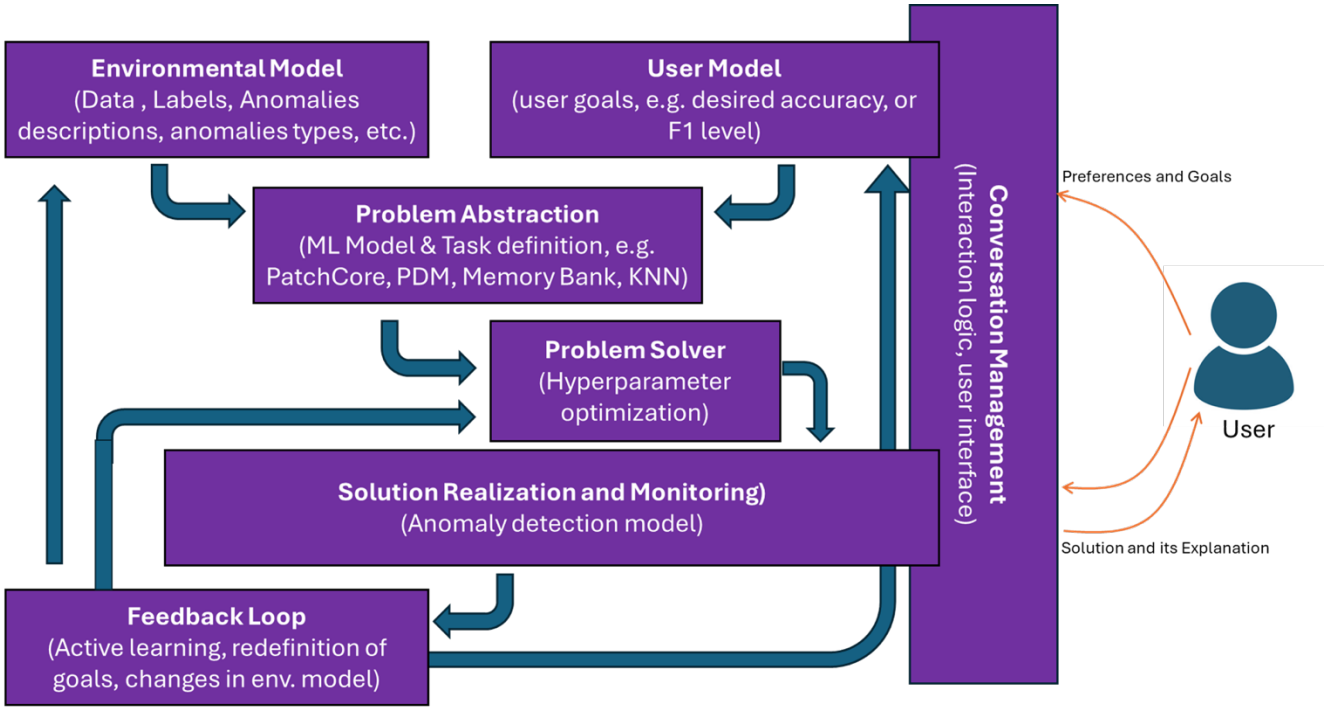


Figure 17 Proditec PEER building blocks mapping to explainable anomaly detection use case

2.3.5.1 Implementation

For this first MVP iteration, we decided with the researchers to use an interactive Figma Mockup that will reproduce the behaviour of the assistant and integrate the solution later when the module is mature enough.

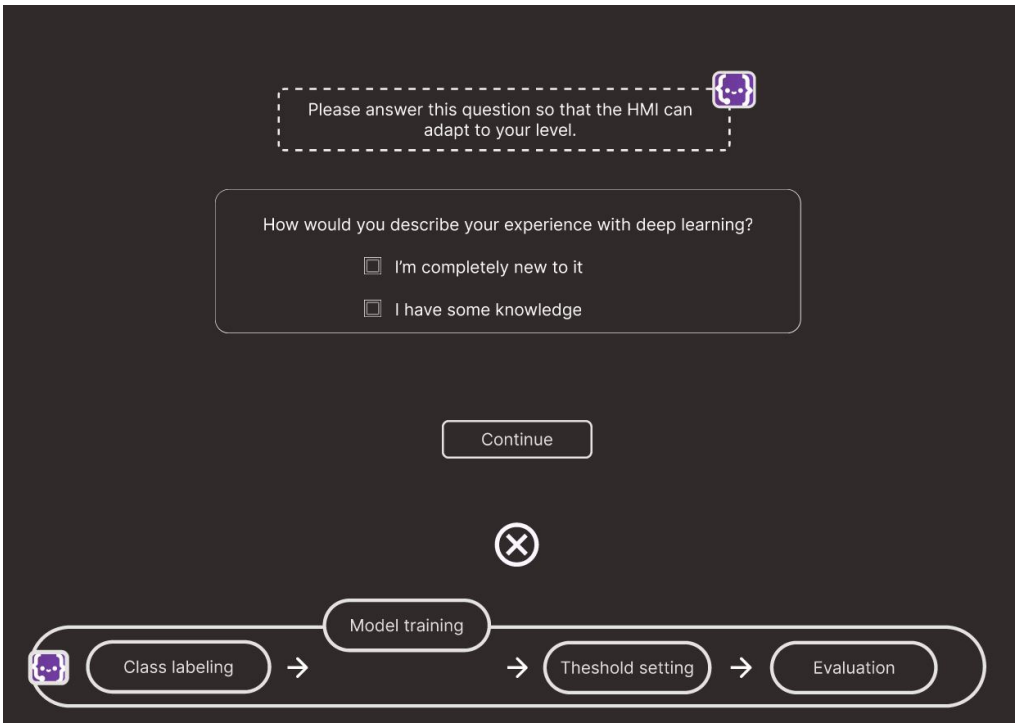


Figure 18 Proditec Trainer Assistant Screen

HMI 3: The trainer assistant collects some information about the operator skills with machine learning to adapt the hyperparameter configuration step to his profile. We use autoML computation to search for the best values in case the operator is new to ML.

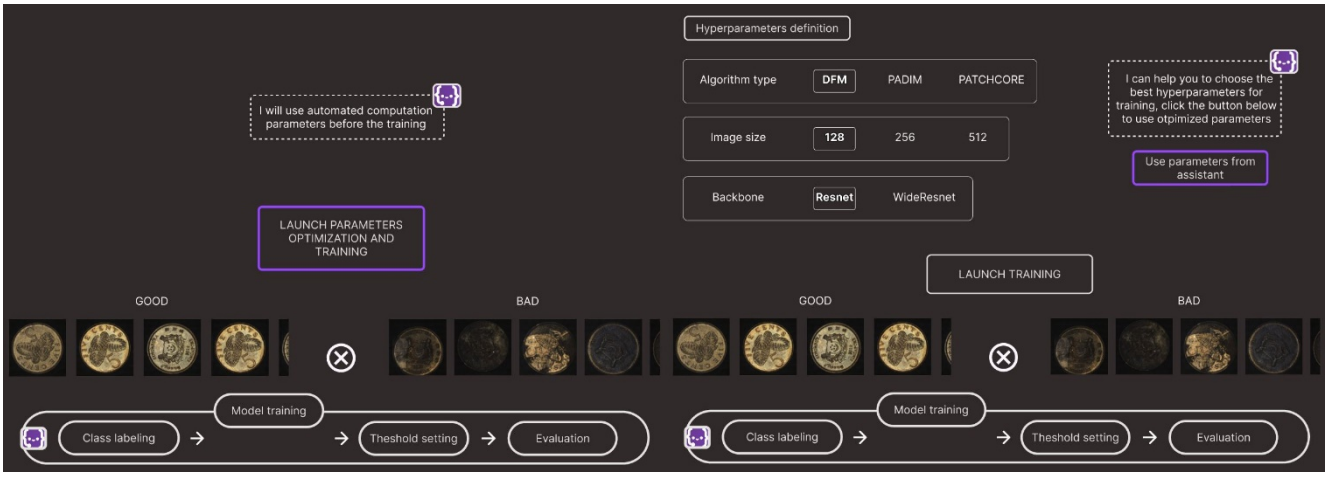


Figure 19 Proditec Hyperparameter Configuration Screen

HMI 4: On the left the screen to proceed the hyperparameters optimization in case the user is not specialist of machine learning. On the right, the hyperparameters manual settings for advanced user, the operators can also choose saved hyperparameters.

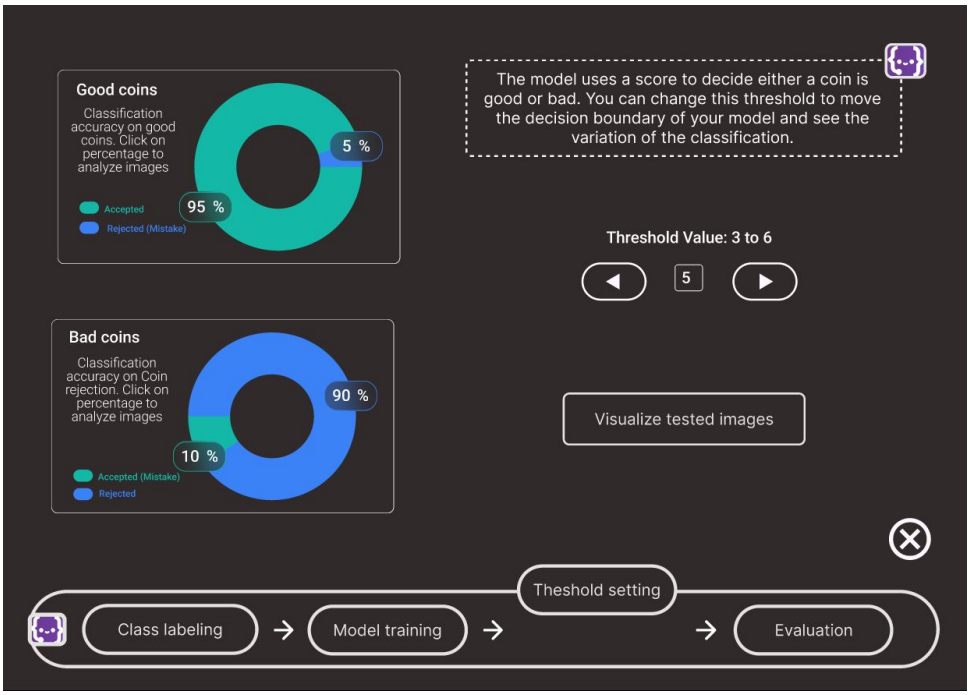


Figure 20 Proditec Threshold Setting Screen

HMI 5: Anomaly detection models generate score during inference. This score represents the deviation of the analyzed image from the dataset he learned. It is therefore necessary to adjust a threshold depending on the sorting requirements of the factory or the cash center. This threshold is mainly used to equilibrate false rejection and false acceptance during production mode.



Figure 21 Proditec Defect Visualization through Heatmaps Screen

HMI 6: This screen helps the user to understand what kind of defects are detected by associating heatmap to each image using SHAP layer. This step helps also to detect model malfunction and aberrant prediction.

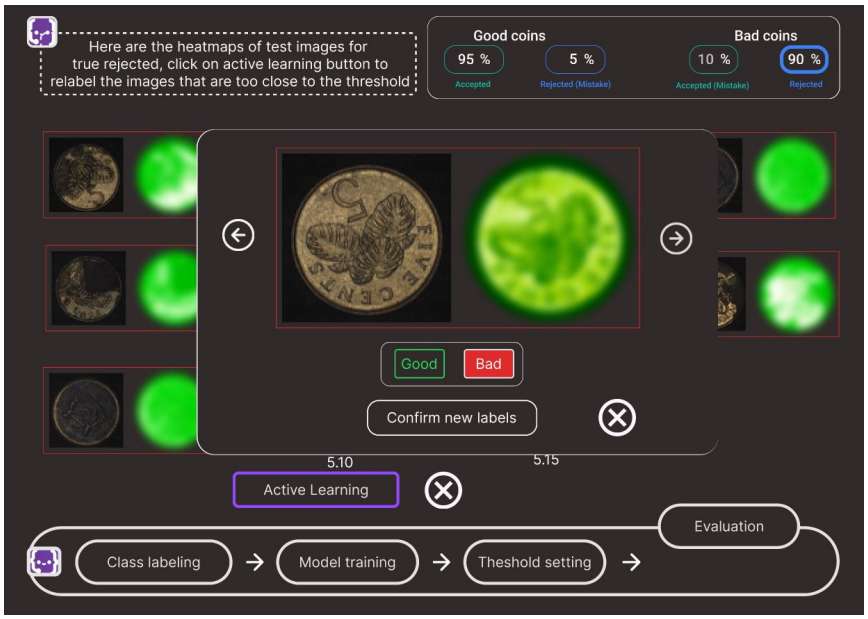


Figure 22 Proitec Active Learning Screen

HMI 7: The active learning module of the assistant will help the user to improve training dataset quality by doing training iterations. The first feature is to relabel the wrong predictions in the test dataset to add them to the training dataset, the second is to reclassify images that have deviation scores too close to the threshold and the third is to add new dataset during training iterations.

2.3.6 Responsibilities

Modules	Modules	Responsible
Semi-supervised labelization	Research on the best model for initial labelization	JU - Szymon Bobek
Explainability layer	Research on how to improve heatmap precision	JU - Szymon Bobek
Active learning	Research on active learning integration for dataset annotation	VUB - Mehrdad Asadi
Frontend	Design mockup of PDT's software	PDT - Princy
Evaluation review	Evaluation of the MVP	CATIE/PDT

2.4 Datacation

2.4.1 Motivation

The Datacation use case is driven by the need to optimize warehouse operations in increasingly complex and dynamic logistics environments. Current methods for configuring warehouse zones and allocating storage can be rigid and may not adapt efficiently to future demands. There is a need for an advanced AI system to assist logistics engineers and systems experts in optimizing the allocation of location types within a flexible digital warehouse configuration, specifically based on future incoming orders. The core motivation is to enhance decision-making through a collaborative Human-AI interaction, ensuring that users remain central to the process, thereby improving efficiency, adaptability, and user trust in the system's recommendations.

2.4.2 Goals

The primary goal for the Datacation MVP is to develop an advanced AI system designed to assist logistics engineers and systems experts in optimizing the allocation of location types within a flexible digital warehouse configuration based on future incoming orders. The system prioritizes a collaborative Human-AI interaction, ensuring that users remain central to the decision-making process.

The specific objectives for this MVP are:
- Develop Mock-ups for the user interface (UI)
- Create a simplified warehouse for initial testing
- Develop a format to provide end-users with possible solutions
- Define the scope of the MVP
- Determine a clear set of objectives and constraints for the system
- Precomputed configurations
- Precomputed explanations/interactions

2.4.3 Limitations

For the first MVP the following are explicitly out of scope:

- Development of a digital twin of the warehouse (this is planned for future iterations).
- Development of a production-ready solution.
- Development of a full-fledged, independent interface (focus is on mock-ups and interaction logic).
- Development of a real-time dialog system (interactions will be based on defined mechanisms for this stage).
- Integrations with warehouse management system and forecasting tools.

- We don't take into account future orders for this MVP, we only focus on the order that causes the error in the WMS.

2.4.4 User Journey

The user journey for the Datacation MVP will involve logistics engineers and system experts interacting with the AI system to optimize warehouse configurations.

1. The user enters the application
2. User input:
    a. The user accesses the system and inputs the following:
        i. The current configuration of location type segregation zones within the warehouse
        ii. The current state of the warehouse
        iii. The details of the incoming order that triggered the error
            1. Note: Future MVPs will support multiple incoming orders
3. Initial analysis:
    a. The AI system simulates the WMS and identifies the reason for the error. It then provides a clear explanation for the issue.
4. Solution generation and explanation:
    a. The engineer starts the optimization. The system generates two optimized layout proposals for the warehouse's location type segregation zones.
        i. A visual layout of the new warehouse configuration
        ii. An explanation of why each solution was chosen based on defined objectives
5. User interaction and feedback:
    a. The user can now interact with the system by:
        i. User provides preferences, constraints, or context information
        ii. User suggests solution
6. Improvement loop:
    a. The process is iterative, allowing for repeated optimization cycles (step 4-5) based on user input and feedback until a good solution is found.
7. Final output:
    a. Once the user approves a solution, the system outputs the optimized configuration of the location type segregation zones. The user can then choose to implement the new configuration in the warehouse.

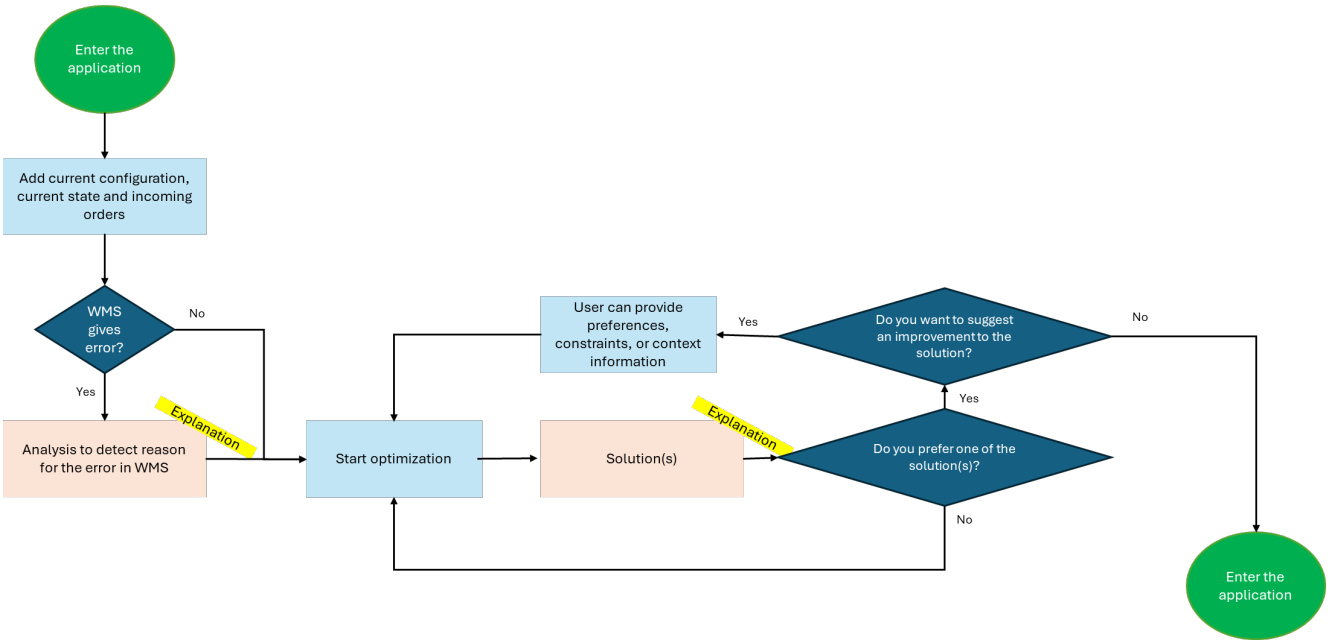


Figure 23 User journey Logistic Engineers of warehouse

2.4.4.1 User flow implementation

For this first MVP iteration, we decided with the researchers to use an interactive Figma mockup to show the engineers how our AI system could look like. The screens below represent one scenario, during the experimental design, multiple scenarios will be tested.

The first step is logging in to the tool, The login page is shown in Figure 24.

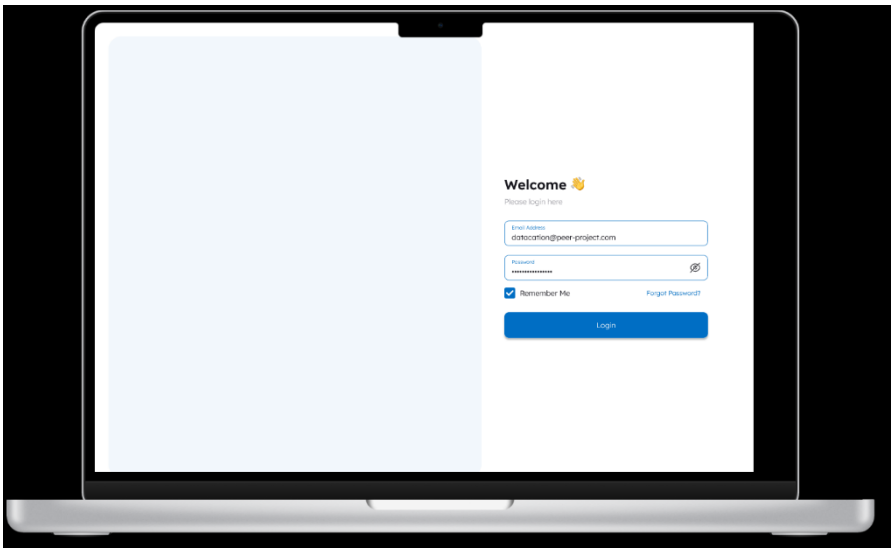


Figure 24 Landing page for login

After logging in, the engineer is directed to the input page (see Figure 25), where they are prompted to upload three documents:

- The current warehouse configuration
- The current state of the warehouse
- The list of orders

Each document has an ‘Upload File’ button, which opens a file explorer so the engineer can select the appropriate file. To ensure clarity, templates will be provided so users know exactly what data is required by the tool.

Once a file is successfully uploaded and validated, the corresponding button turns green, providing immediate visual feedback. After all files are uploaded, the engineer can proceed to configure the tool by clicking the ‘Configure’ button. The system will then confirm whether the configuration was successful (Figure 25).

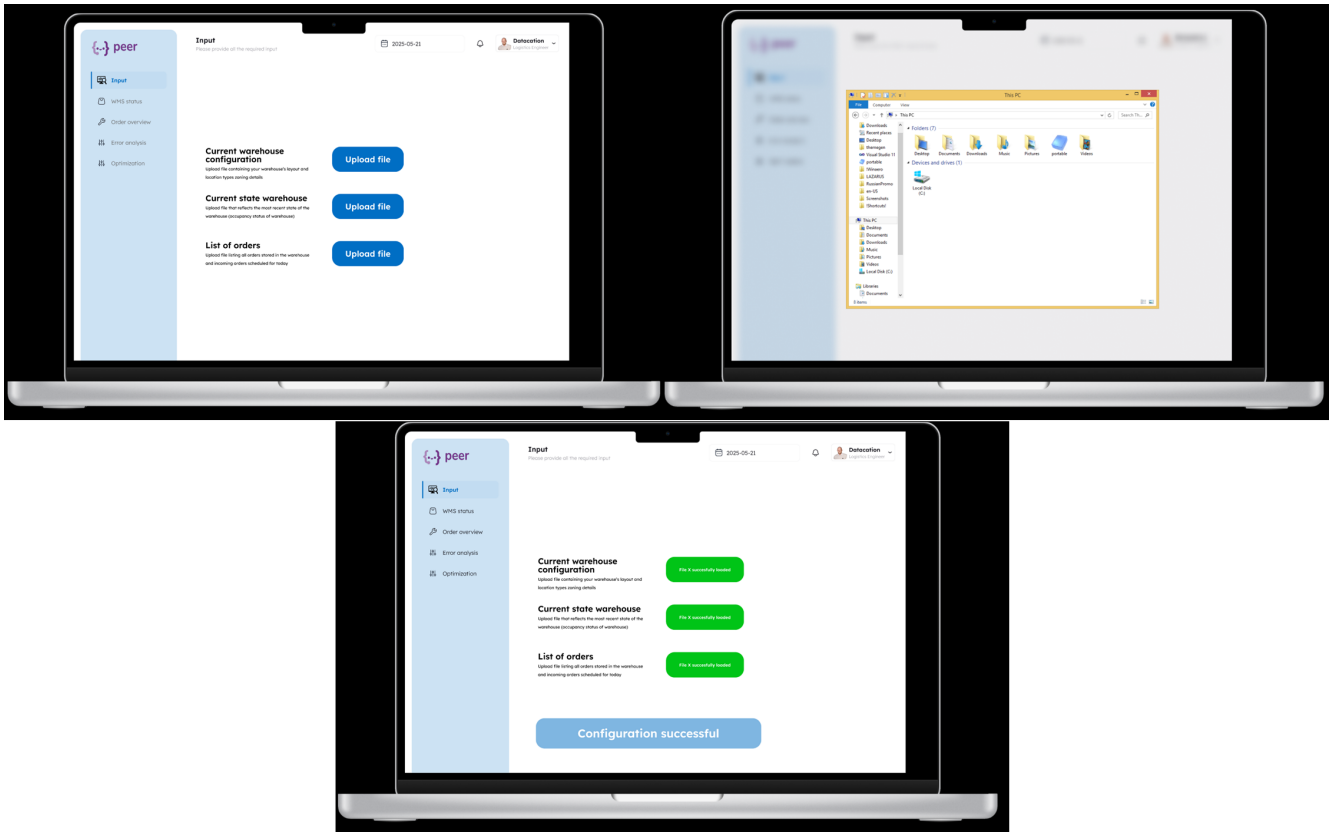


Figure 25 Uploader pages after login for configuration parameters

After the configuration, the user is free to select one of the following pages:

- WMS status --> to see a current overview of the warehouse configuration and state
- Order overview --> to see a list of order currently stored or incoming orders of the day
- Error analysis --> to get an analysis of which orders cause an error, including explanation
- Optimization --> to start with the optimization process

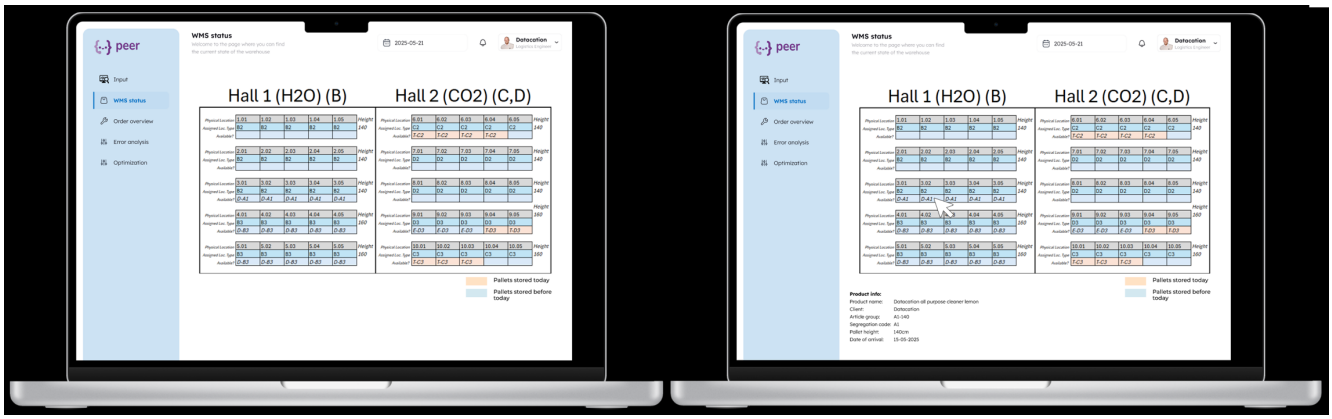


Figure 26 Check the current status of the warehouse

For now, we assume that the engineer would like to first review the current status of the WMS to gain deeper insights into the warehouse's current state. Figure 26 presents a visualization of this page. To view detailed information about a specific pallet, the engineer can simply hover over the pallet, which will display the corresponding product information.

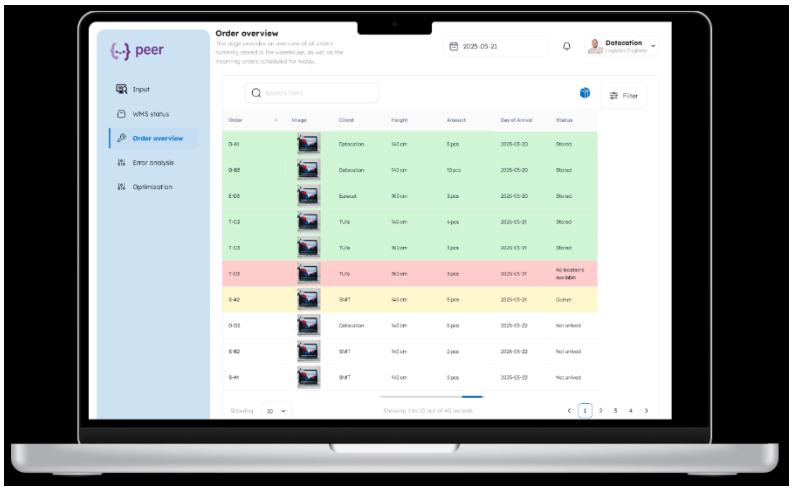


Figure 27 Order overview

On the **order overview page** (Figure 27), engineers are presented with a summary of all orders, including details such as order name, client, height, quantity, arrival date, and status.

- **Green** indicates orders that are already stored in the WMS.
- **Red** highlights orders that cannot be stored in the WMS due to the current configuration.
- **Yellow** represents orders that are in the queue.

In addition to the screenshots included, there are also screens that offer to add a filter.

To investigate an issue, the engineer can select an order marked in red. This action leads to the error analysis page, where an explanation is provided for why the order cannot currently be stored in the warehouse.

The error analysis page (Figure 28) provides a comprehensive overview, including the current state of the warehouse, detailed information about the specific order causing the error, and a clear explanation of why one or multiple pallets from the incoming order cannot be stored. Additionally, this page offers an option to initiate the optimization process.

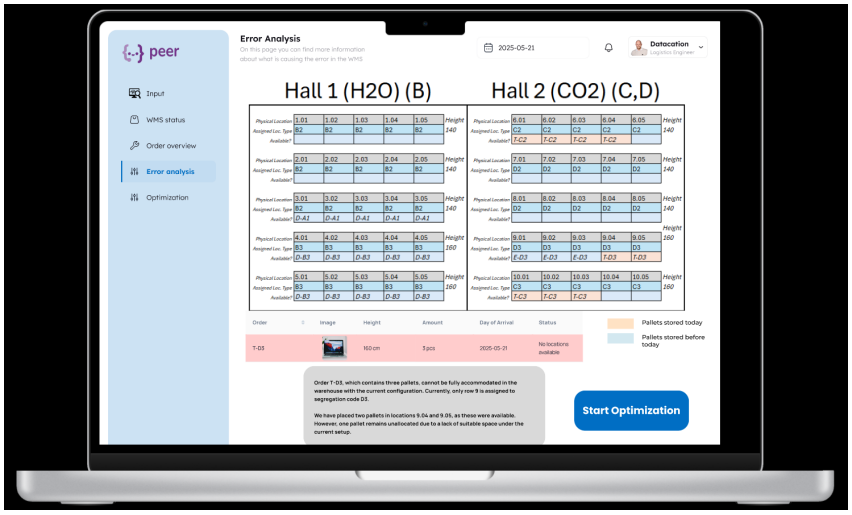


Figure 28 Error analysis page

During the optimization process, a loading screen is displayed (Figure 29). Once the optimization is complete, the engineer is presented with two proposed solutions. The results page consists of multiple subpages, see bar below. The first subpage (Figure 29) displays the four defined objectives along with the scores for each solution, allowing for an immediate side-by-side comparison. For further insight, the engineer can click the explanation button to view more detailed information about either solution.

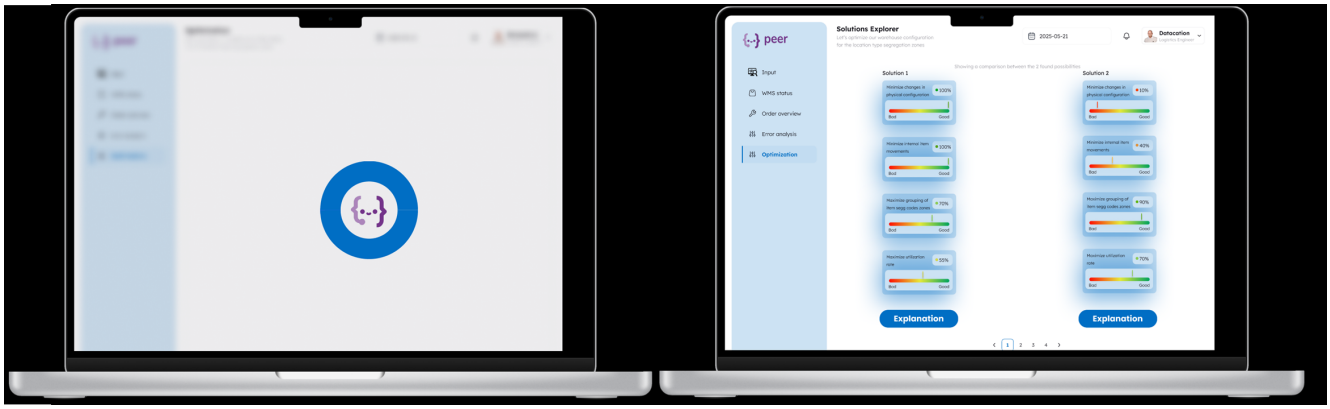


Figure 29 Loading screen and solution comparison

Subpages 2 and 3 present visualizations of the proposed solutions. As with previous page, there is an option to access further explanations via the explanation button. Figure 30 illustrates both the visualization screen and the explanation screen.

In the visualization, any remaining pallets that previously caused the error are highlighted in yellow. Configuration adjustments are displayed in bold, while the previous location type segg code is also

shown to provide clear insight into the changes made. The explanation screen offers contextual information using clear language and includes a comparison of the objectives.



Figure 30 Solutions explorer with additional explanation

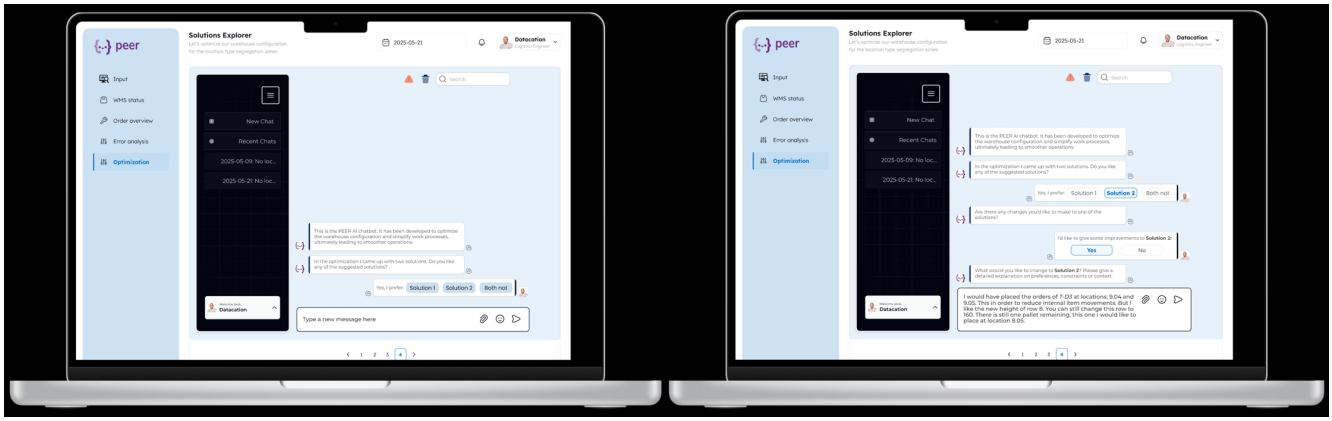


Figure 31 Solutions explorer with the chatbot

Subpage 4 (Figure 31) features a chatbot interface, where the AI tool begins by asking the engineer which of the proposed solutions he/she prefers.

The engineer can respond with one of three options: Solution 1, Solution 2, or Neither.

If the engineer selects a solution (the “happy flow”), the chatbot then asks whether they would like to make any adjustments to the chosen solution (Figure 31). Otherwise, the optimization starts all over again. If the engineer answers “no,” the new warehouse configuration is immediately exported as an Excel file (Figure 33), concluding the user flow.

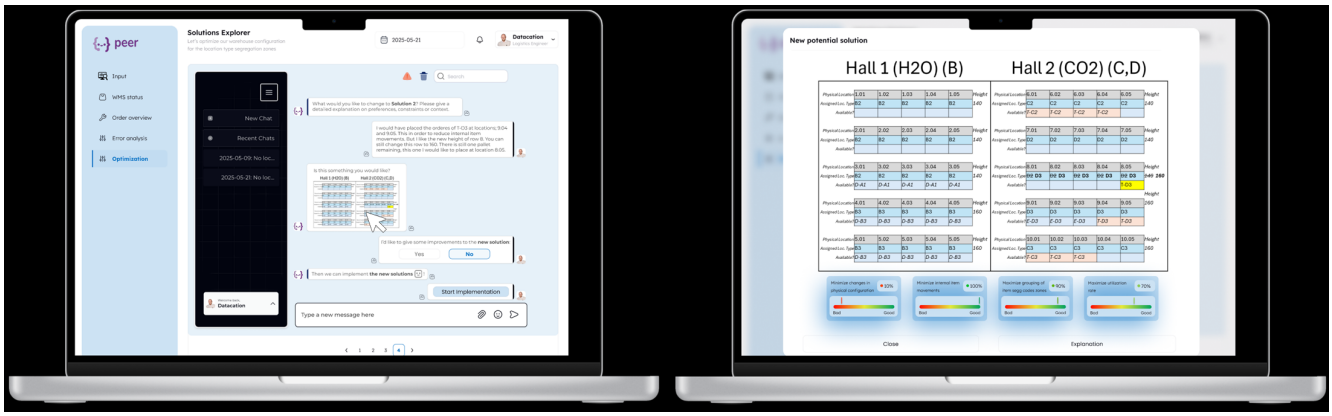


Figure 32 Solutions explorer with the chatbot

If the engineer answers “yes,” they can provide textual feedback regarding their preferences, constraints, or any additional context (Figure 31). Based on this input, the AI generates a new solution, which is then visualized within the chat (Figure 32). It is possible to select the adjusted solution to get more insights via the score of the objectives or the explanation button (Figure 32). The engineer can continue to provide improvements if desired. When satisfied, the updated configuration can be exported as an Excel file (Figure 33), marking the end of the user journey.

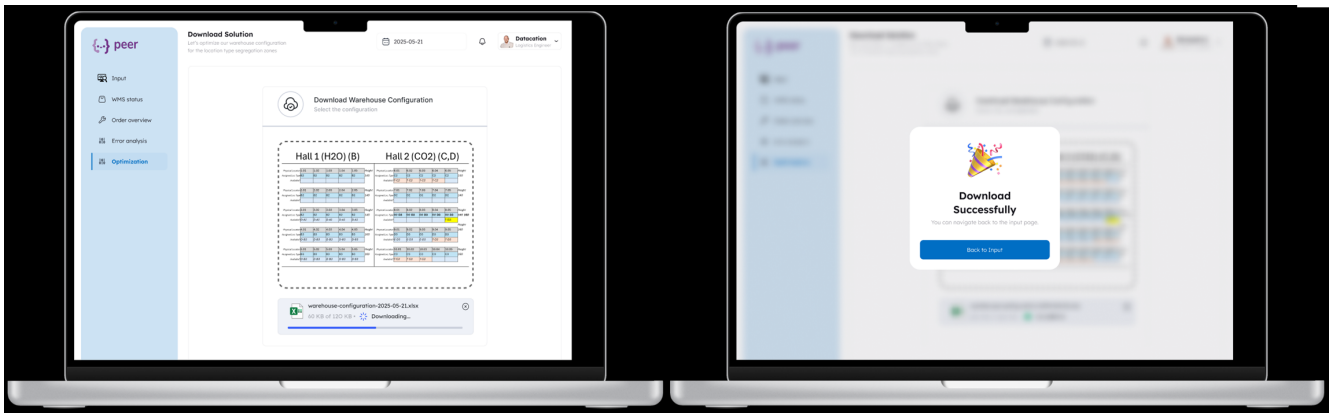


Figure 33 Downloading the approved solution

2.4.4.2 User-AI Interactions

The interaction between the user and the AI is central to this MVP. Key aspects include:

- Collaborative Configuration: The AI does not replace the human expert but assists them. Users provide critical inputs (current state, forecasts, preferences, constraints).
- Explainable Suggestions: The AI will not just provide a configuration but also explain why a particular configuration is suggested. This is crucial for building trust and enabling informed decisions. The "AI explanation" module, potentially LLM-based, will facilitate this.
- Feedback Loop: Users can provide feedback on the AI's suggestions. This feedback is vital for understanding user acceptance and for the potential future learning of the AI system.

- Questioning Capability: The system aims to allow users to ask further questions to clarify the AI's suggestions, moving towards a more conversational interaction.
- Interface: Interaction will initially be through UI mock-ups and defined data formats (e.g., Excel). An "Interface/Abstraction Module (LLM-based)" is envisioned to manage these interactions.

2.4.4.3 Socio Technical Requirements

Question	Answer
What parts/features of the system do you want to test?	Mock-ups for UI, the simplified warehouse simulation, the core AI allocation module, the explanation mechanism, the format for providing suggestions (e.g., Excel), and the overall Human-AI interaction flow, including how users provide preferences and feedback.
Why do you want/plan to test it? What do you hope to learn (testing goal)?	To learn how to effectively model warehouse processes, material flows, and decision points. To understand how to model the Human-AI interaction, including user preferences for collaboration with the AI. To assess user acceptance and trust in the AI system.
With whom do you want to do the test (e.g., end-users)?	Logistics engineers and Systems Experts (as direct and indirect end-users respectively).
When do you want/plan to do the test (timing)?	MVP #1 is to be delivered and tested by 31.06.2025. Key activities like User-AI interaction specification and Figma design are planned for April-June 2025.
What tools do you have available to perform the test (e.g., Figma, wireframes, nothing)?	Mock-ups for UI (Figma designs by DAT), a simplified simulation of the warehouse, and a format for suggestions (e.g., Excel).
What will be (technically) working in the MVP (e.g., AI module)?	The core AI allocation module (decision-making module) suggesting warehouse configurations, a simplified warehouse simulation, the mechanism for generating explanations, and a defined format for presenting suggestions to the user.
What needs to be simulated? What cannot be simulated?	A simplified simulation of the warehouse, orders, allocation decisions, and their consequences will be developed. A complete digital twin or a production-ready, real-time dynamic interface will not be part of MVP1.

2.4.5 Evaluation Journey

The evaluation of MVP #1 will focus on understanding:

- Warehouse Modelling: How effectively the simplified simulation and AI models capture the essential warehouse processes, flows, and decision-making logic.
- Human-AI Interaction:
    - User preferences regarding how they want to collaborate with the AI system.
    - The clarity and usefulness of the AI's explanations.
    - The ease of providing preferences, constraints, and feedback.
    - The effectiveness of the chosen format (e.g., Excel, UI mock-ups) for presenting suggestions.
- User Acceptance and Trust: Initial assessment of user acceptance of the AI's suggestions and their trust in the system's capabilities.
- System Utility: Whether the AI's suggestions are perceived as valuable and potentially leading to better warehouse configurations.
- Scope Validation: Feedback on the defined scope and identification of critical areas for future development.

The evaluation will be guided by questions such as the effectiveness of the tested features, what is learned about modelling and interaction, and overall user acceptance.

2.4.6 Design of the MVP

The MVP design involves several key modules and data inputs, aiming for a collaborative system where the AI assists human experts.

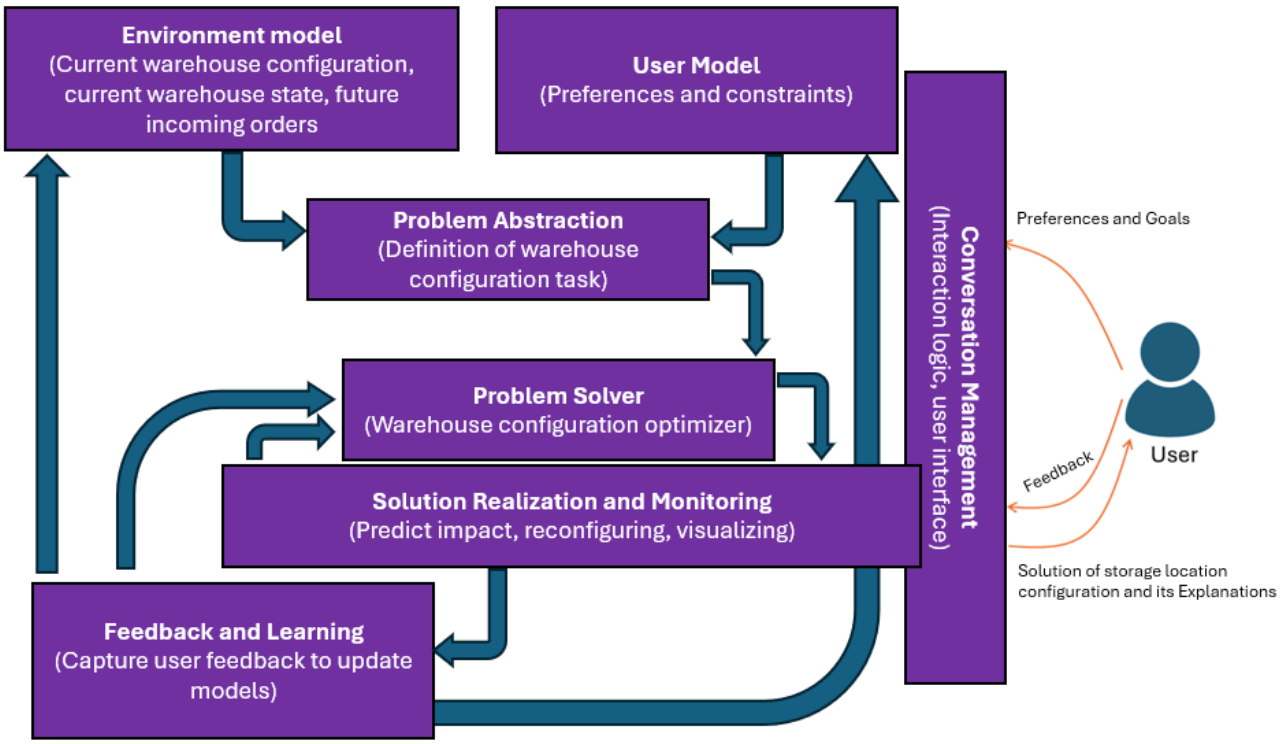
Core Components

- Data Inputs:
    - Current State Warehouse (WMS): Provided by DAT. Contains the current configuration of the warehouse and available space. Essential for the AI to understand the starting point.
    - Forecast: Provided by DAT. Contains forecasts of future orders, necessary for the AI to make decisions on adjusting the online warehouse configuration proactively.
    - Users' Preferences, Parameters / Constraints / Weights: Input from logistic engineers/system experts. This allows users to guide the AI's decision-making process according to specific operational needs or goals.

- AI Modules:
    - Interface/Abstraction Module (LLM-based): Developed by EUT. This module will manage the interaction between the human user and the AI system, including handling inputs, presenting outputs, and facilitating explanations.
    - Warehouse Simulation: Developed by DAT, with CU leading MVP AI aspects with all AI partners. A simplified representation of the warehouse, orders, allocation decisions, and their consequences. Used for testing and evaluating potential configurations.
    - Decision-Making Module: Developed by TU/e, with CU leading MVP AI aspects with all AI partners. This is the core AI allocation module. For a given warehouse state and order forecast (and potentially future orders in later iterations), it will generate 1-n promising solutions (e.g., using multi-objective optimization) based on inputs from the human-AI interface.

- AI Explanation, Users' Feedback, Storage Location Configuration (XAI aspects): Handled by TU/e, with CU and all AI partners. This cluster of functionalities focuses on making the AI's decisions transparent and interactive. It includes generating explanations for solution quality, comparing solutions, and processing user feedback to potentially refine suggestions or understand preferences better. The output is the suggested storage location configuration.

- Users:
  - Logistic Engineers (Direct end-user)
  - System Experts (Indirect End-user)



2.4.6.1 Design choice due to technology challenges

Given the complexity of a real-world warehouse, we decided to develop a simplified version for our initial design. As a result, integration with a full warehouse management system is not yet feasible. Instead, we have created a design prototype with the following characteristics:

- A fixed current configuration and state of the warehouse
- No live prototype available for MVP1
- Preliminary (prefix) solutions in place

We developed design concepts based on this simplified warehouse model, focusing on predefined objectives and preferences. However, we acknowledge that end-users may have additional requirements, particularly regarding the configuration of location type segregation zones.

2.4.7 Responsibilities

The following table outlines the responsibilities for the different modules and tasks within the Datacation MVP:

Category	Module/Task	Explanation	Responsible	Due Date
Data	Current state warehouse (WMS)	The current configuration of the warehouse in the WMS. Needed to see available space.	DAT	01-03-2025
Data	Forecast	Forecast of future orders. Needed to decide on how to adjust the online warehouse configuration.	DAT	31-12-2024
AI	Interface/Abstraction	The interface to provide the interaction between the human and AI.	EUT	TBD
AI	Warehouse simulation	A simple representation of the warehouse, orders, allocation decisions, and their consequences for the warehouse.	DAT, CU (leading MVP AI) together with all AI partners	TBD
AI	Decision-making	For a given warehouse and order, create 1-n promising solutions, e.g., with multi-obj. optimization taking input from human-AI interface.	TU/e, CU together with all AI partners	TBD
X(AI)	AI explanation, users' feedback, Users' preferences, Storage location configuration	For the human-AI interface - generate explanations for solution quality, comparisons between solutions (later also projecting into future).	TU/e, CU together with all AI partners	TBD

3. Conclusions

This report details the first release of the Minimum Viable Product (MVP) for the PEER collaborative AI assistant, showcasing its instantiation across four key use cases: Sonae (Retail), City of Amsterdam (Smart & Inclusive City), Proditec (Quality Control), and Datacation (Logistics Optimization).

The development and implementation of these MVPs are core activities within WP5, contributing to the overall goal of integrating and validating the AI assistant in real-world scenarios.

As such, the MVP presented here will need to be evaluated by WP4 to assess its acceptability through user studies and gather insights on system usability and acceptance.

The AI depth of the MVP, including planning algorithms, XAI, user modelling, and the human-AI conversation loop, is described in detail in Deliverable 3.1, which defines the PEER building blocks and modular architecture.

PEER will focus on how to systematically put the user at the centre of the entire AI design, development, deployment, and evaluation pipeline, allowing for truly mixed human-AI initiatives on complex sequential decision-making problems. The central idea is to enable a two-way communication flow with enhanced feedback loops between users and AI, leading to improved human-AI collaboration, mutual learning and reasoning, and thus increased user trust and acceptance. As an interdisciplinary project between social sciences and artificial intelligence, PEER will facilitate novel ways of engagement by end-users with AI in the design phase; will create novel AI planning methods for sequential settings which support bidirectional conversation and collaboration between users and AI; will develop an AI acceptance index for the evaluation of AI systems from a human-centric perspective; and will conduct an integration and evaluation of these novel approaches in several real-world use cases.






























This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101120406.