TALON

Autonomous and Self-organized Artificial Intelligent Orchestrator for a Greener Industry 4.0

Deliverable

D5.3 Pilot Specific TALON Platform Setup, Operation, Continuous Integration & Maintenance Report

Actual submission date: 04/11/2024

Funded by the European Union

Project Number:	101070181		
Project Acronym:	TALON		
Project Title:	Autonomous and Self-organized Artificial Intelligent Orchestrator for a Greener Industry 4.0		
Start date:	October 1st, 2022	**Duration:**	36 months
D5.3 Pilot Specific TALON Platform Setup, Operation, Continuous Integration & Maintenance Report			
Work Package:	WP5		
Lead partner:	INTRA		
Author(s):	Magda Foti (INTRA)		
Reviewers:	Konstantinos Kyranou (SID); Vasileios Argyriou, Vladislav Li (KU)		
Due date:	31/10/2024		
Deliverable Type:	R	**Dissemination Level:**	PU
Version number:	1.0		

Revision History

Version	Date	Author	Description
0.1	09/09/2024	INTRA	Table of Contents release
0.2	01/10/2024	INTRA	Contributions in Section 1
0.3	14/10/2024	PROBO, FACTOR, KU, CERTH	Contributions in Sections 3 & 4
0.4	16/10/2024	INTRA	Version ready for review
0.5	22/10/2024	KU, SID	Review
0.6	25/10/2024	PROBO, FACTOR	Contributions in reviewers' comments
0.7	25/10/2024	INTRA	Sent for internal quality and content review
0.8	30/10/2024	INTRA	Final version; addressed quality assurance check comments
1.0	01/11/2024	ENG	Final coordinator review before submission

Table of Contents

List of Figures

List of Tables

Definitions and acronyms

AI	*Artificial Intelligence*
APIs	*Application Programming Interfaces*
AR	*Augmented Reality*
CA	*Consortium Agreement*
CI	*Continuous Integration*
CD	*Continuous Deployment (or Continuous Delivery)*
DoA	*Description of Action*
EC	*European Commission*
EU	*European Union*
EC2	*Amazon Elastic Compute Cloud*
E2C	*Edge-to-cloud*
FL	*Federated Learning*
GA	*Grant Agreement*
JSON	*JavaScript Object Notation*
KPI	*Key Performance Indicator*
MS	*Milestone*
NG-SDN	*Next Generation – Software Defined Network*
PaaS	*Platform as a Service*
SLOs	*Service Level Objectives*
TrL	*Trust Level*
UATV	*Automatic unnamed arial and terrestrial vehicle*
UC	*Use Case*
UI	*User Interface*
QoS	*Quality of Service*
XAI	*Explainable Artificial Intelligence*
VR	*Virtual Reality*
WP	*Work Package*

Disclaimer

This document has been produced in the context of TALON Project. The TALON project is part of the European Community's Horizon Europe Program for research and development and is as such funded by the European Commission. All information in this document is provided 'as is' and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability. For the avoidance of all doubts, the European Commission has no liability with respect to this document, which is merely representing the authors' view.

Executive Summary

The vision of TALON is to design and develop next-generation industrial systems in terms of performance, adaptation, explainability, trustworthiness and transparency. TALON aims *to pave the way for* the next industrial revolution by developing a fully automated AI architecture capable of bringing intelligence near the edge in a flexible, adaptable, explainable, energy- and data-efficient manner. In this direction, TALON researches and develops a rich set of components that enable intelligent, energy-efficient, trusted, human-centred, AI-based orchestration of resources across the cloud-edge continuum. Most importantly, the project integrates and demonstrates these components in several real-life use cases and demonstrators. The implementation of such demonstrators is a challenging task, as it requires the integration of diverse and technologically heterogeneous components, *necessitating structured planning* of the integration work.

In this context, the present “Deliverable D5.3 – Pilot Specific TALON Platform Setup, Operation, Continuous Integration & Maintenance Report” documents TALON’s demonstrators regarding set-up and system integration process. It provides a pilot specific platform set-up and the integration schedule at component level for each demonstrator, along with an early evaluation of the KPIs defined. Specifically, D5.3 provides (i) a detailed report on the integration status per pilot and the timeline at which the TALON platform setup will be completed and (ii) the first results on the evaluation of the KPIs as they were defined in “Deliverable 2.1 - Use Case, KPIs, Requirements, Specification, Slices & Technology Enablers Definition Report”. In the final deliverable of WP5, “Deliverable D5.4 - Final TALON Platform Setup, Operation, Continuous Integration & Maintenance Report” a final and complete evaluation of the KPIs will be presented, providing additional validation and evidence how the KPIs have been achieved.

1 Introduction

1.1 Objective of the Deliverable

The primary objective of deliverable D5.3 Pilot Specific TALON Platform Setup, Operation, Continuous Integration & Maintenance Report is to document TALON's demonstrators regarding set-up and system integration process. The current deliverable reports on the integration activities in each demonstration site at component level providing integration status and the future integration plans. In addition, each demonstrator provides an early version of the evaluation results, reporting on the quantification of the pilot specific KPIs. The aim of this report is to initiate the TALON demonstration activities and ensure a successful implementation of the integration, validation and demonstration activities of the project.

1.2 Relation to other Work Packages

This is the third deliverable of "WP5 - Integration, Validation and Demonstration" series of deliverables. The work reported in this document is part of demonstrators' tasks, namely: Task 5.3: Automatic UATV Coordination, Task 5.4: I5.0 Automation & Planning, Task 5.5: AR/VR for Training & Maintenance and Task 5.6: Human Robot Collaboration. These tasks, and the output of deliverable D5.3, rely on input from D2.1, D3.1, D3.2, D3.3 and D4.1. In addition, the outputs of this deliverable will be used as part of T3.1: Overall Architecture & Platform Design; T3.2: Enabling Common NG-SDN & Distributed Intelligence Functionalities; T3.3: Zero-Touch AI-Orchestrator; T3.4: AI-based Resource Coordination Through Computation Offloading & Social Aware Caching; and T4.4: Security & Privacy Blockchain Mechanisms.

Completing D5.3 marks the attainment of Milestone 6 (MS6) – Availability of TALON Demonstrators.

The operations of Task 5.3, Task 5.4, Task 5.5 and Task 5.6 will continue after the submission of D5.3 until the end of the project (M36) in order to support the demonstration of the TALON results. The final evaluation of the TALON results and the demonstration activities will be presented in D5.4: Final TALON Platform Setup, Operation, Continuous Integration & Maintenance Report (M36).

1.3 Structure of the Document

This document is structured in 4 Chapters:

- Chapter 1 provides an overall introduction to the deliverable objectives and their positioning within the project.
- Chapter 2, entitled " Pilot Specific TALON Platform Setup" provides a detailed status update on the platform setup per demonstrator together with the future integration planning. Chapter 2 specifies the TALON architectural components that are used in each pilot site and provides the integration status at component level.
- Chapter 3, entitled "Early Evaluation Results", provides a first quantification of the KPIs as they were defined per pilot site.
- Finally, Chapter 4 concludes the deliverable along with the outlook.

2 Pilot Specific TALON Platform Setup

The next paragraphs present the requirements that are relevant with each one of the TALON's pilot sites, with each requirement defining the relevant architectural components. The following tables also provide the deployment status of each component in the pilot sites together with a time plan for the next deployment activities.

2.1 Automatic UATV Coordination

	REQ	STATUS	Comments
Access and Security Requirements	Authentication and Authorisation	Not yet	To be Deployed by end of M27
AI-fueled Orchestration	Service Modelling (aka "Configurations) and Enactment	Deployed	
	AI Swarm Orchestration	In progress	To be Deployed by end of M30
	Resource Allocation and Deployment	Deployed	
	Definition, Customisation and Monitoring of Metrics	Deployed	
	Data Monitoring, Collection and Aggregation	Deployed	
	AI Model Training and SLOs Optimisation	Deployed	
AI Cognition	Hybrid and Optimised Learning	In progress	To be Deployed by end of M28
	Data Operations	Deployed	
	Visualisation Dashboard	Not yet	To be Deployed by end of M30

2.2 I5.0 Automation & Planning

	REQ	STATUS	Comments
Access and Security	Authentication and Authorisation	Deployed	
	DLTs for Securing AI/ML models weights (DLTs for Security and Privacy)	In progress	To be Deployed be end of M26
	Anomaly Detection	In progress	To be Deployed be end of M26
AI-fuelled Orchestration	Service Modelling (aka "Configurations) and Enactment	In progress	To be Deployed be end of M27
	Definition, Customisation and Monitoring of Metrics	In progress	To be Deployed be end of M27
	Data Monitoring, Collection and Aggregation	In progress	To be Deployed be end of M27
AI Co	Self-healing and Self-correcting	In progress	To be Deployed be end of M26

	Hybrid and Optimised Learning	In progress	To be Deployed be end of M26
	Data Operations	In progress	To be Deployed be end of M27
	Digital Twins	In progress	To be Deployed be end of M25
	XAI, Monitoring and Reporting	In progress	To be Deployed be end of M27
	Data Lifecycle Management	In progress	To be Deployed be end of M27
	Visualisation Dashboard	In progress	To be Deployed be end of M27

2.3 AR/VR for Training & Maintenance

	REQ	STATUS	Comments
Access and Security	Authentication and Authorisation	In Progress	To be Deployed be end of M25
	Data Anonymisation	In Progress	To be Deployed by end of M25
AI-fuelled Orchestration Requirement	Service Modelling (aka “Configurations) and Enactment	In Progress	To be Deployed by end of M27
	Resource Allocation and Deployment	In Progress	To be Deployed by end of M27
	Definition, Customisation and Monitoring of Metrics	In Progress	To be Deployed by end of M27
	Data Monitoring, Collection and Aggregation	In Progress	To be Deployed by end of M27
	AI Model Training and SLOs Optimisation	In Progress	To be Deployed by end of M27
AI Cognition Requirements	Hybrid and Optimised Learning	In Progress	To be Deployed by end of M26
	AI Capabilities and Transfer Learning	In Progress	To be Deployed by end of M26
	Digital Twins	Deployed	
	Data Lifecycle Management	In Progress	To be Deployed by end of M26
	Visualisation Dashboard	In Progress	To be Deployed by end of M27

2.4 Human Robot Collaboration

	REQ	STATUS	Comments
Access and Security	Authentication and Authorisation	Deployed	
	Data Anonymisation	In progress	To be Deployed by end of M25
AI-fuelled Orchestr	Service Modelling (aka “Configurations) and Enactment	Not Yet	To be Deployed by end of M27

	Resource Allocation and Deployment	In progress	To be Deployed by end of M27
	Definition, Customisation and Monitoring of Metrics	In progress	To be Deployed by end of M27
	Data Monitoring, Collection and Aggregation	In progress	To be Deployed by end of M27
	AI Model Training and SLOs Optimisation	In Progress	To be Deployed by end of M27
AI Cognition	Data Operations	In progress	To be Deployed by end of M26
	Data Lifecycle Management	In Progress	To be Deployed by end of M26
	Visualisation Dashboard	In Progress	To be Deployed by end of M27

3 Early Evaluation Results

This section provides an early view on the evaluation results per pilot site. The KPIs, as defined in Deliverable "D2.1: Use Case, KPIs, Requirements, Specification, Slices & Technology Enablers Definition Report" are presented together with the results of the first TALON field tests.

3.1 Automatic UATV Coordination

The current values of the KPIs defined below represent the actual results tested in the field to date. The KPIs selected are tailored to the evaluation of the complex scenario and addresses statistics coming from the drone use and internal processing.

In more detail the current KPIs are as follow:

- KPI_01 – The network probing capabilities is a challenging KPI. To achieve the results a lot of QoS was necessary.
- KPI_02 – The UATV to node communication is a very sophisticated process and due to the complexity, the optimization techniques are elegant both in software and hardware. The current results are near to 10ms which is pretty close to our target!
- KPI_03 – This is all about compression of the data and smart use of hash tables. We are trying to find a more intelligent way to lower the 16 Bytes to 4 Bytes.
- KPI_04 – This KPI has been achieved as in general we have on average 20 Watt/h.
- KPI_05 – The AI recognition is pretty much achieved. An effort to go further than 95% is mostly unrealistic.
- KPI_06 – This KPU can be considered achieved but requires further investigation for improvements.

The core KPIs are driven from Energy Efficiency and Data Efficiency. Those two correlate and are intertwined due to the nature of the use case. In fact, the process we followed is the following:

- Run the software part in the simulator
- Run all drones in the simulator
- Set acceptable parameters and change them to try to align to optimal values
- Catalog the results
- Re-iterate and reset until we have the best approximations
- Test the results in the field based on the simulated values
- Refactor code and acceptable parameters on the drones
- Retest in the field
- All settings now are set close to the target KPIs

Table 1: Use Case 1 Key Performance Indicators' Early Evaluation

Key performance indicators

ID	Name	Description	Reference to mentioned use case objectives	How will be measured / assessed?	Baseline Value	Target Value	Early Results
KPI_01	Latency reduction	Decrease response time to < 150ms	Latency reduction	Network probing tool, measured by $\lvert T2 - T1 \rvert$	50ms	24ms	25ms
KPI_02	UATV-to-Node communication latency	>90% decrease in UATV-to-Node feed forwarding latency	UATV-to-Node communication latency	Network probing tool, measured by $\lvert T2 - T1 \rvert$	20ms	≤10ms	8ms
KPI_03	EE and Data Efficiency	>80% reduction in transferred data/size	Energy Efficiency; and Data Efficiency	Data rate measure tool, measured by $\frac{\Delta size_{before}}{\Delta size_{after}}$	A few KB	A few bytes	16B
KPI_04	EE on operated flights	>30% energy conservation on operated flights	Energy Efficiency	On-chip VCC/A measurements, measured by $\lvert E2 - E1 \rvert$	30+ Watt	About 20 Watt	20Wattt
KPI_05	AI – Valid Detections	>Faster and more accurate recognitions mean less flight time	Accuracy	Measured based on actual objects on screen	80%t	95%	~94.7%
KPI_06	Valid Actions (Triggers) from AI	Triggered actions for predefined scenarios (operational efficiency)	Data & Energy Efficiency	Match against original defined actions	190	100	~102

3.2 I5.0 Automation & Planning

The current values of the KPIs defined below reflect the actual values in the shopfloor. The KPIs selected are the typical KPIs used in a manufacturing company to evaluate the efficiency of the processes and the whole factory.

The main indicator in a manufacturing company is the OEE (Overall Equipment Effectiveness), which identifies the percentage of manufacturing time that is truly productive. The OEE is calculated by multiplying Availability, Quality and Performance. Specifically in Factor, Quality and Performance are close to 100%, the Availability being the main factor that needs improvement to reach 100% OEE.

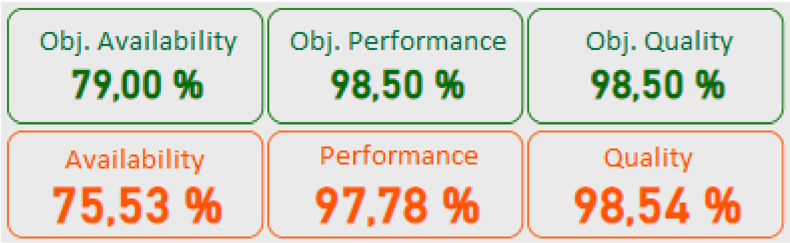

Figure 1. Availability, Performance and Quality of the Nakamura2 in the last period.

The availability of the machine measures the time that the equipment is available for production. It is calculated by dividing the actual running time by the planned production time.

The actual running time is the total time the equipment is operational, including downtime for planned maintenance, unplanned breakdowns, and idle time. The planned production time is the scheduled operating time for the equipment.

The factors affecting the availability of the machine are:

- Unplanned downtime. Breakdowns, equipment failures and unexpected maintenance.
- Planned downtime: Scheduled maintenance, cleaning and inspections.
- Idle time: Periods when the equipment is not running due to lach of work or material.

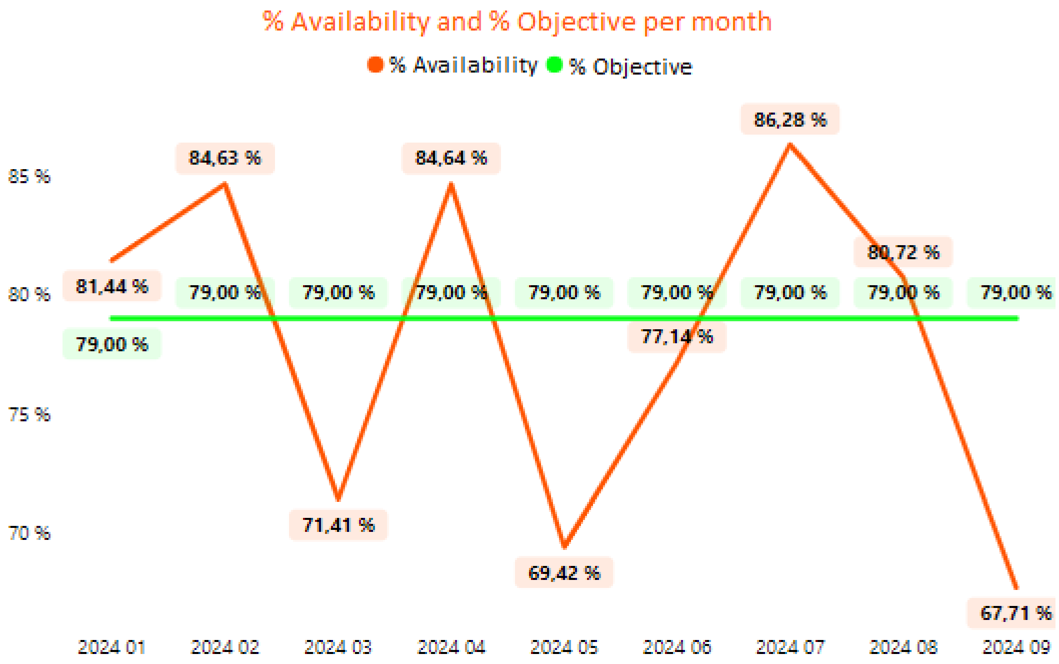

Figure 2. Availability of the Nakamura2 during 2024

The Quality ratio in a manufacturing environment measures the number of good quality products produced compared to the total number of products produced. It is calculated by dividing the number of good quality units produced by the total number of units produced.

Factors affecting quality are:

- Process control. The ability to maintain consistent product quality.
- Inspection. The effectiveness of quality control measures.

- Operator skills. The skill and training of operators.
- Raw material quality. The quality of the materials used in production.

Depending on the value of the parts or the material, an improvement of 1% in the quality ratio can represent around 10.000 € monthly in a company.

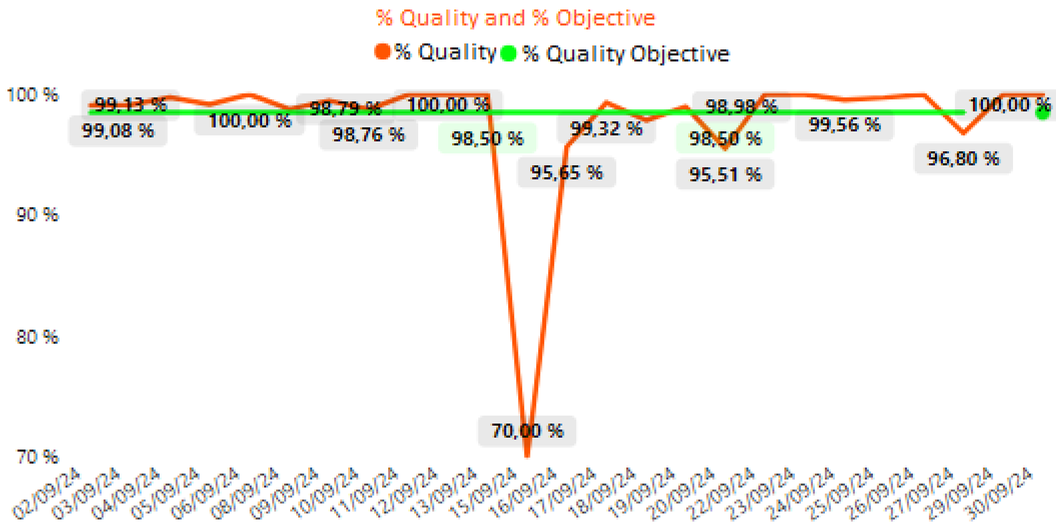

Figure 3. Quality Ratio of the Nakamura2 during September 2024

By these components and their interrelationships, manufacturers can identify specific areas for improvement and implement strategies to increase overall equipment effectiveness.

Table 2: Use Case 2 Key Performance Indicators' Early Evaluation

Key performance indicators							
ID	**Name**	**Description**	**Reference to mentioned use case objectives**	**How will be measured / assessed?**	**Baseline Value**	**Target Value**	**Early Results**
KPI_07	Quality ratio increase	Quality ratio is the measurement of the poorly manufactured parts within the manufacturing process. Is the result of the division of the well-produced parts over the total	Quality ratio increase by increasing the number of manufactured parts	This KPI is measured daily in the company, measured by $(\frac{V2-V1}{\lvert V1 \rvert}X100)$	98%	99% (This improvement in the manufacturing environment represents around 10.000 € / month for the company)	98 %

		manufactured parts.					
KPI_08	Reduction of scrap parts	The scrap is the number of parts that cannot be accepted due to quality issues. The optimisation of the process will reduce this scrap by reducing the parts that are manufactured with wrong dimensions.	Reduction of scrap parts	% of wrong parts, measured by $(\frac{V2-V1}{\lvert V1 \rvert}X100)$	2,2%	2% (This improvement in the manufacturing environment represents around 2.000 € / month for the company)	2,2, %
KPI_09	Lower environmental footprint	The optimisation of the process and the reduction of the scrap will reduce the environmental footprint, since the machine will be working for manufacturing only good parts. If the process is not optimised, it will produce some wrong parts, that leads in a high environmental footprint.	Lower environmental footprint	ISO 14001 Carbon footprint index	Under calculation (Dec.2024)	Under calculation (Dec.2024)	Under calculation (Dec. 2024)
KPI_10	Increase in availability	Availability is the relation of the time that the machine is manufacturing good parts and the time that the machine should have been manufacturing good parts. Every time that the machine is stopped for any reason, it is not manufacturing good parts,	Increase in availability	This KPI is measured daily in the company. It will be measured by $(\frac{V2-V1}{\lvert V1 \rvert}X100)$	78%	79% (This improvement in the manufacturing environment represents that the whole factory would work 7,2 more hours per month, which represents 11.000 €/month)	78 %

		which implies a loss in productivity.					
KPI_11	Decrease investigation efforts	Decrease the need of manual investigation in the events of loss of signal in the optical links, limiting the manual investigation to the erroneously classified events	Decrease maintenance time	Compute the time taken by the ML classifier to infer the class of a time-series associated with the optical fibre link fault, measured by ($\frac{T2-T1}{\lvert T1 \rvert}X100$ and $\lvert T2-T1 \rvert$)	Focusing on the time it takes to carry out the root cause analysis, in an optical network equipped with integrated Optical Time-Domain Reflectometer (OTDR), the investigation time carried out by expert network engineers can be estimated, from experiences conducted by research laboratories. This is estimated to be approximately a couple of hours compared to the response time of an ML classifier of <1sec	In case of a correct ML-driven inference, decrease the time it approximately takes is estimated to be <1sec	
KPI_12	Increase optical fault analysis reliability	Increase the reliability in the overall fault analysis process with the support of explanations from the machine learning model	Improve investigation quality	Measured as Cross Lipschitz Extreme Value for nEtwork Robustness (CLEVER) for XAI reliability	Currently, reliability is measured by observation. New measures are being introduced, as XAI is an emerging topic.	[l_2<0.7, l_∞<0.05], Least Likely <1.5, Random Target<1.5, and Top-2 Target<0.98	

3.3 AR/VR for Training & Maintenance

This section presents the early evaluation results of using AR/VR technologies for training and maintenance. The key performance indicators were measured through simulated training sessions and real-world maintenance tasks, focusing on improving task completion times, reducing errors, and increasing user engagement. The data was collected through user feedback, task analysis, and system logs to evaluate the performance and usability of the AR/VR tools.

The early experiments demonstrate an estimated response time based on 10 minutes of monitored data communication between the AR maintenance application and the backend solution. The response time for each request, considering the given data, can be calculated by adding the transmission time and the latency. Based on a total data transfer over the total duration of the monitoring (5 requests per second), the average transmission time per request is approximately 7.58 milliseconds. Adding the round-trip latency of 14 milliseconds, the total response time for each request is approximately 21.58 milliseconds listed in Table 5.

At present, the assessment of the effectiveness of AI-human collaboration has yet to be undertaken, as depicted in Table 5. We are currently in the preparatory phase of this process, focusing on finalising the methodology and collecting the requisite data. We expect to commence the evaluation in the near future and the findings will be disseminated as soon as they become available, in accordance with our established project timeline. This thorough evaluation aims to provide valuable insights into the dynamics of human and AI interaction, enhancing our understanding of collaborative effectiveness in this context.

The measurement of the AR-to-node point-of-view (POV) transmission latency reduction was conducted by measuring the time between the data being rendered on the screen of the support personnel in frames per second (fps). The implementation of the new system has resulted in a reduction in AR-to-node POV transmission latency, achieving a decrease of 90% from the baseline value of 2 seconds (T1) to just 200 milliseconds (T2). This improvement calculated using the formula $T2 - T1$, where T1 represents the initial latency and T2 the post-implementation latency, demonstrates a latency reduction of 1.8 seconds. Further, the difference is converted into a percentage for evaluation using the formula $\frac{T2 - T1}{T1} \times 100$. The percentage decrease, quantified at 90%, ensures faster data transmission between the AR device and the node. The initial and post-implementation latency were measured by recording the time taken to send, receive, and render a frame on the screen, expressed in fps. The initial latency was measured at 0.5 fps, while the post-implementation latency improved to 5 fps

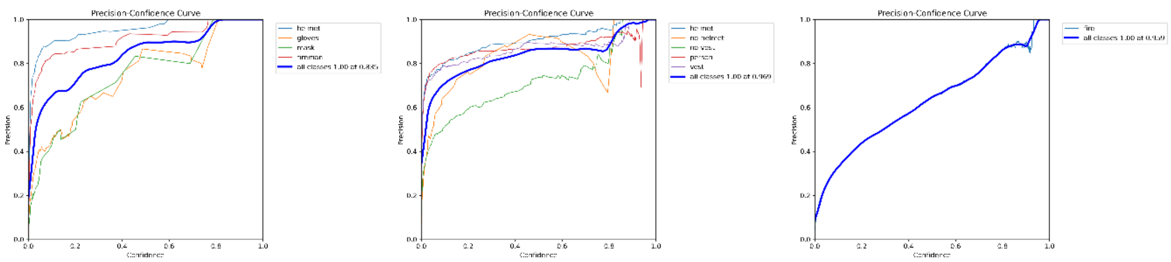

Figure 4.The diagrams are showing the precision-confidence metric that visualises the performance of initial recognition accuracy of object detection model employed in the pilot representing the initial results.

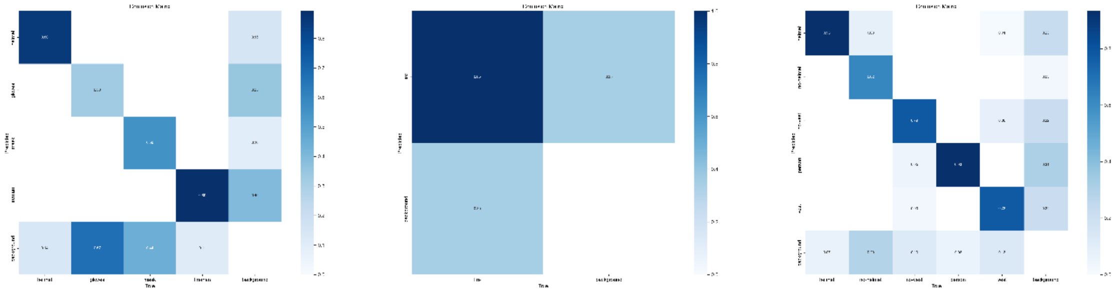

Figure 5. The diagrams are demonstrating the confusion matrices for construction safety, fire, and personal protective equipment (from left to right) representing the initial results.

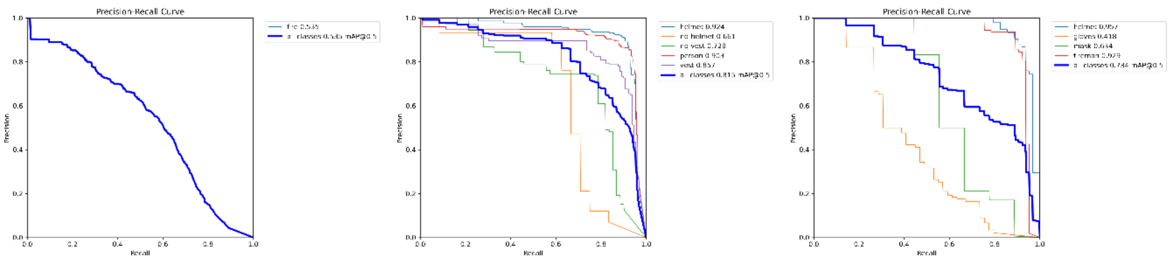

Figure 6. The graphs are representing precision-recall curves the visualise the performance of the models based on construction safety, fire, and personal protective equipment datasets (from left to right) providing the initial results.

The recognition accuracy in object detection can be articulated as mean average precision (mAP), as it highlights the performance of selected predictors for the environment recognition. The preliminary findings illustrated in Figure 7 not only surpass the established baseline, illustrated in Figure 4, but also approach the target accuracy with commendable precision. These elevated performance metrics are further substantiated by additional auxiliary metrics frequently employed in the domain. For instance, the confusion matrices shown in Figure 8 provide detailed insights into classification outcomes suggesting similar positive trend when compared with Figure 5, while the mean average precision results presented in Figure 6 and Figure 9 reinforce the robustness of the model's performance. Together, these metrics offer a comprehensive evaluation of the model's efficacy in object detection tasks.

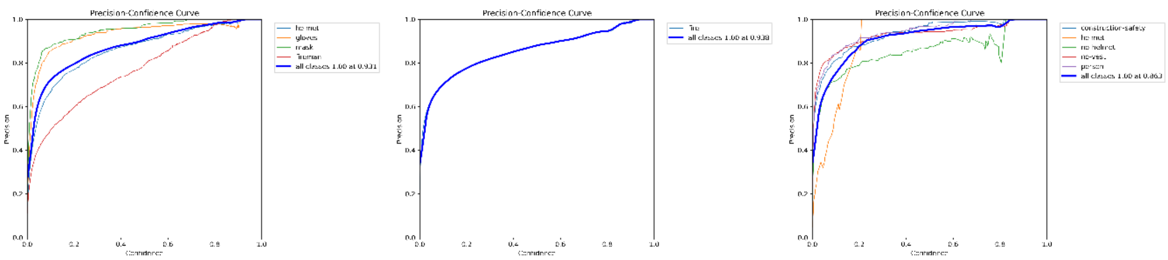

Figure 7. The diagrams are showing the precision-confidence metric that visualises the performance of post-implementation recognition accuracy of object detection model employed in the pilot demonstrating the early results.

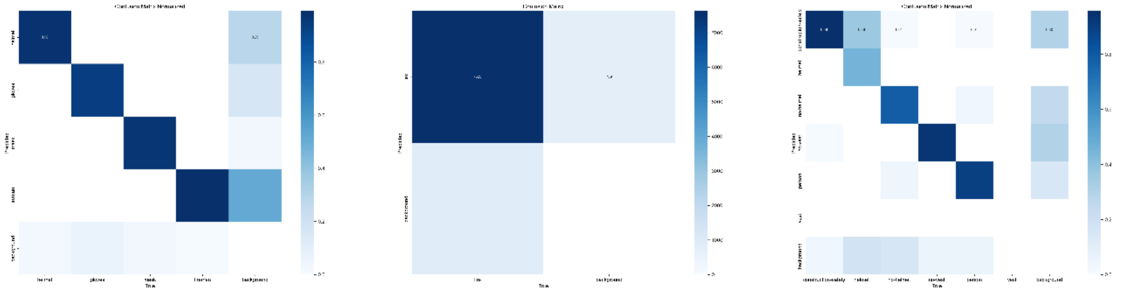

Figure 8. The diagrams illustrate the confusion matrices for construction safety, fire, and personal protective equipment (from left to right), highlighting early high-accuracy results.

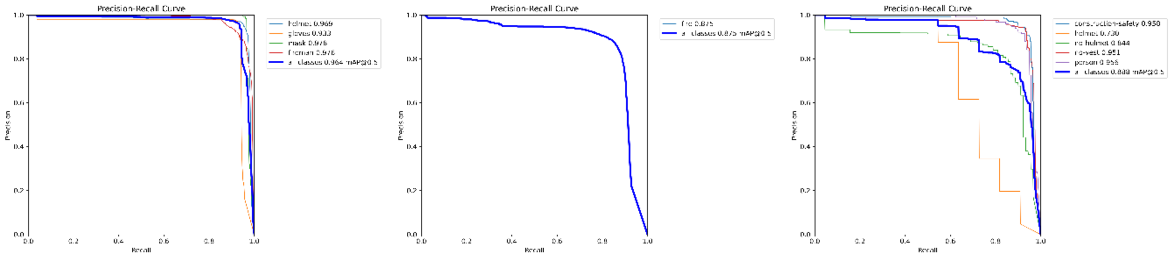

Figure 9. These graphs depict precision-recall curves for construction safety, fire, and personal protective equipment (from left to right), indicating a strong trend towards early high-accuracy results.

The results for the environment/object recognition accuracy are collected through a series of structured tests and performance evaluations. The process involves comparing the system's recognition outputs against a predefined set of known environments, i.e. ground truth. Each test consists of presenting the system with various images, and the system's ability to correctly identify the objects within images is measured. The correct identification rate is calculated by dividing the number of correctly recognised environments by the total number of ground-truth. It calculates the average precision (AP) for each class by considering the precision-recall curve, then averages the AP values across all classes. Regular benchmarking is performed to ensure consistency, and performance evaluations are conducted under different conditions. Data from these tests are aggregated to determine the overall accuracy, and any misclassifications are analysed for further system refinement. The results are updated periodically as the system is re-tested with improved data and algorithms.

Table 3. The table below lists the model performance results, represented by the mAP metric, highlighting high recognition accuracy.

mean Average Precision (%, mAP)	Initial Results	Early Results
Construction Safety	81.5%	86.2%
Fire	53.5%	87.5%
PPE	73.4%	96.4%
Average	69.47%	90.03%

Table 3 displays the initial and early results of the recognition accuracy. Using the formula $\left(\frac{I}{E}-1\right)\times 100$, where *I* represents the initial results, and *E* represents the early results, the model trained on the construction safety dataset shows a positive change of 21.97%, indicating that the modifications to the initial model were successful. Similar improvements are observed for the fire and PPE datasets at 12.02% and 23.86%, respectively. The current system, initially at the performance indicated in Table 3, underwent several improvements. Data augmentation and expansion were applied to increase the diversity of the training dataset by incorporating variations in lighting, angles, and environmental conditions, allowing the system to better adapt to real-world scenarios. The recognition model was fine-tuned by optimising its architecture, adjusting hyperparameters, and implementing transfer learning from pre-trained models to enhance performance. Comprehensive testing was conducted regularly, using a wide range of data to ensure consistent benchmarking and performance evaluation. These steps successfully improved the system's performance, aligning it with the target value mentioned in Table 5.

The personally identifiable information (PII) underwent a thorough audit using OpenCV, during which substantial sets of images were randomly selected from three distinct datasets, each pertaining to different environments: fire safety, construction safety, and the use of personal protective equipment (PPE). The face anonymisation algorithm was applied to each selected image, followed by a subsequent deanonymisation algorithm to assess the robustness of the anonymisation process. This methodology achieved a 100% rate of non-identifiability concerning personal data for the detected faces only. The audit unequivocally confirmed that all facial data had been effectively anonymised, thereby ensuring the complete safeguarding of workers' sensitive and personally identifiable information. This process demonstrates the effectiveness of the anonymisation techniques employed and reinforces the commitment to maintaining the privacy and confidentiality of individuals in these critical environments.

Table 4. A table demonstrating an approximate number of images that were sampled for the re-identification purposes to conform with the KPI_17.

Name	Number of Images
Fire	~1000
Construction Safety	~3500
Personal Protective Equipment	~300

Table 5: Use Case 3 Key Performance Indicators' Early Evaluation

Key performance indicators							
ID	**Name**	**Description**	**Reference to mentioned use case objectives**	**How will be measured / assessed?**	**Baseline Value**	**Target Value**	**Early Results**
KPI_01	Latency reduction	Decrease response time to < 20ms. This is related to the AR maintenance scenario focusing on	Latency time reduction	The response time will be measured and assessed by utilising performance monitoring tools to track network	50ms	20ms	21.58ms

		the AR to edge communication		latency and response times in real-time, implementing logging within the AR application to record precise timestamp at request initiation and response reception, and analysing this data to ensure the response time reaches below 20ms, measured by ($\lvert T2 - T1 \rvert$)			
KPI_13	AI-human collaboration effectiveness	>50% AI-human collaboration effectiveness. Increase of the human satisfaction during the AR-maintenance scenario that is powered by AI models. And improved user experience and interaction during training activities	AI-human collaboration effectiveness	The AI-human collaboration effectiveness will be measured through a combination of surveys and feedback forms to gauge human satisfaction and user experience during the AR-maintenance scenario and training activities. Additionally, performance metrics will be collected to evaluate the efficiency and accuracy of tasks completed through AI-human collaboration, aiming for over 50% effectiveness. User	2.0 average on 5-point Likert system	3.0+ average on 5-point Likert system	N/A

				interaction data and qualitative assessments will further validate improvements in user experience and interaction, measured by a 5-point Likert system			
KPI_14	AR-to-node AR POV latency reduction	>90% decrease in AR-to-node POV transmission latency	Decrease in AR-to-node AR point-of-view (POV) transmission latency	The >90% decrease in AR-to-node POV transmission latency will be measured by comparing the latency metrics before and after implementing the new system. Continuous monitoring tools will record the time taken for data to transmit from the AR device to the node. Baseline latency values will be established prior to the new system's deployment, and subsequent latency measurements will be taken to ensure they reflect a decrease of over 90%. Regular performance reports and	2s	200ms	200ms

				real-time analytics will be used to validate the latency reduction such as in the case of data required for the training with and without few-shot learning, measured by $(\|T2 - T1\|)$			
KPI_15	Increased reusability	Training attendance rate > 95%	Increased reusability and training attendance rate	The training attendance rate of >95% will be measured by tracking participant attendance records for each training session. Attendance logs will be maintained and reviewed to calculate the percentage of attendees relative to the total number of invited participants. The assessment will be conducted regularly to ensure the attendance rate consistently meets or exceeds 95%, and reports will be generated to document compliance with this metric.	1/20 participants	19/20 participants	N/A

KPI_16	Increased environment recognition	>90% accuracy through increased environment recognition	Increased gesture / environment recognition	The >90% accuracy in environment recognition will be measured by comparing the system's recognition outputs against a predefined set of known environments. The system's recognition accuracy will be evaluated through a series of tests, where the correct identification rate is calculated by dividing the number of correctly recognised objects by the total number of test objects. Regular performance evaluations and benchmarking against these known environments will ensure the accuracy consistently meets or exceeds the 90% threshold, measured by $(\frac{V2-V1}{\|V1\|}X100)$	70% accuracy	>90% accuracy	*90.03%*
KPI_17	PII preservation	100% PII preservation via facial, numerical, and textual data anonymisation	Preserve workers' sensitive and personal identifiable information	The 100% PII preservation through facial, numerical, and textual data anonymisation will be	n/a	100%	100%

				measured by conducting audits on the anonymised data aiming to reidentify the related removed information. These evaluations will involve cross-referencing the anonymised data against the detection algorithms that were used initially to ensure no personally identifiable information (PII) can be retrieved, measured by $(\frac{V2-V1}{\lvert V1 \rvert}X100)$			

3.4 Human Robot Collaboration

The scenario for the UC4 concerns the use of the TALON system in a factory for the intelligent, efficient and timely identification of workers who should be wearing personal protective equipment (P.P.E.) such as helmets, vests, and other safety gear. The previous situation for the above process was completely empirical with the supervision of the premises being done by a safety officer. Instead, now a drone will be used to supervise the site alongside the TALON system to identify them. Several KPIs have been defined within this scenario related to the performance of the individual TALON systems that will be involved in the scenario. Table 6 below presents in detail all the KPIs defined in the context of this scenario.

Table 6: Use Case 4 Key Performance Indicators' Early Evaluation

Key performance indicators							
ID	**Name**	**Description**	**Reference to mentioned use case objectives**	**How will be measured / assessed?**	**Baseline Value**	**Target Value**	**Early Results**
KPI_16	Increased object / environment recognition	>90% accuracy through increased	Increased object recognition	Experiments for accuracy between previous and optimised AI	85%	>90%	90%

		object recognition		Models, measured as Delta function to calculate the improvement $\lvert Acc2 - Acc1\rvert$			
KPI_17	Preservation of anonymisation	100% preservation of anonymisation via facial, numerical, and textual data anonymisation	Preserve workers' sensitive and personal identifiable information	The anonymized image is re-processed through the same face detection model used. The absence of detected faces confirms the preservation of the anonymization.	AS-IS no anonymisation is applicable	100% (numerical, textual and imagery)	
KPI_18	AI-to-AI communication	>70% reduction in AI-to-AI communication latency	Reduction in AI-to-AI communication latency	Latency measured before and after TALON's AI benchmarking and optimisation, measured as $(\lvert T2 - T1\rvert)$	AS-IS time for AI-to-AI communication is <1sec	TO-BE time for AI-to-AI communication will be <0.25ms	
KPI_19	Human vs Robot Inspection time	>30% decrease in inspection time	Robot production efficiency	Timer, measured as $(\lvert \mathrm{T2} - \mathrm{T1}\rvert)$	120 sec (person walking)	30 sec (drone flying)	For 4 places Person time: 20-25 min Drone time: 4-6 min
KPI_20	Robust AI models	>80% in environment recognition and augmentation accuracy	Robust computer vision models	Experiments for accuracy between previous and optimised AI Models, measured as Delta function to calculate the	60%	>80%	75%

				improvement $\|Acc2 - Acc1\|$			
KPI_21	Protective equipment readiness	>30% decrease in protective equipment readiness	Robust computer vision models	Timer, measured as $\|T2 - T1\|$	240 sec (time to raise an alert)	70 sec	For 4 places Person time: 180 min Drone time: 120 min
KPI_22	Personnel safety	>70% increase in the personnel safety	Multimodal object detection	Measured by means of AS-IS and TO-BE difference, expressed by AI Model accuracy as $\|Acc2 - Acc1\|$	AS-IS: no AI / ML method is currently used; only visual personnel inspection	>70% personnel safety increase, expressed by a robust object classifier / detector (i.e., accuracy, f1, precision, recall)	

More specifically, as the pilot provider, CERTH has undertaken 2 KPIs that are more user-oriented. In particular, KPI_19 Human vs Robot Inspection time, where it measures the time needed to supervise all the sites before and after the use of the TALON to highlight the benefit that results. Additionally, KPI_21 Protective equipment readiness which measures the time it takes to raise an alert for a worker not wearing the necessary protective equipment, again before and after using the TALON system.

Regarding the KPIs, some challenges were initially faced. The main one was the intermittent and consistent measurement of the specific KPIs for the current situation, where based on UC4 scenario the process is practically done by a human. In this case it was necessary to ensure repeatability in the measurements. The second struggle was the lack of a suitable space in the premises of CERTH, since the flight of a drone in the working environment was not allowed.

For the above reasons the initial values given in the KPIs, are estimates based on running the scenario in a large hall. To better dissect numbers and provide what is needed for the scenario, the working environment of the logistic company, VANOS S.A., was used. Information about the techniques applied in the factory, for the supervision of their premises were also investigated, as well as the ways these techniques could be applied to the UC4 scenario, thus overcoming initial constrains.

Detailed information was collected about the procedures employees follow in terms of the supervision of the factory premises for security compliance. More specifically, the security officer conducts 4 rounds per day at different time slots (e.g. 9:30, 11:30, 14:00 and 16:00) in all the premises concerned and completes/signs a document containing the time of the surveillance and details of how many workers are in the premises and whether they must wear any personal protective equipment and which equipment in particular. Below, Figure 10, is an indicative document to be completed by the security officer:

<table>
<tr><td colspan="7">Location: Accounting records</td></tr>
<tr><td colspan="4" rowspan="2">Space requirements (PPE):</td><td>Helmet</td><td>Vest</td><td>Glasse s</td></tr>
<tr><td>✔/x</td><td>✔/x</td><td>✔/x</td></tr>
<tr><td colspan="4">Number of Mandatory PPE on site</td><td colspan="3"></td></tr>
<tr><td>Route no</td><td>Entry time</td><td>Exit time</td><td>Number of Employees</td><td></td><td></td><td></td></tr>
<tr><td rowspan="3">1</td><td rowspan="3">9:30</td><td rowspan="3">9:50</td><td>Employee 1</td><td></td><td></td><td></td></tr>
<tr><td>Employee 2</td><td></td><td></td><td></td></tr>
<tr><td>Employee 3</td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td rowspan="3">2</td><td rowspan="3">11:30</td><td rowspan="3">12:50</td><td>Employee 1</td><td></td><td></td><td></td></tr>
<tr><td>Employee 2</td><td></td><td></td><td></td></tr>
<tr><td>Employee 3</td><td></td><td></td><td></td></tr>
<tr><td colspan="7">.....</td></tr>
</table>

Date Signature

Figure 10: Security officer's document

Therefore, the link between the above process and the document to the UC4 KPIs is clear. KPI_19 is directly linked to the time it takes for the security officer to supervise an area, while KPI_21 is linked to the rounds the security officer does in relation to the drone where it can do more rounds, so the time to raise an alert will be reduced. Subsequently, to ensure representativeness and repeatability in our results, the security officer was asked to collect and provide copies of the documents he/she signs for each site within a week with a typical workload. Hence, based on these documents an average of the times for the UC4 KPI's for the situation before the use of the TALON system will be obtained. Then, after TALON is adopted by the company, the corresponding times will be measured and compared to the initial values, before the use of TALON system. Since the process before was conducted by a human, at the end of the project using a drone and the TALON system, significant improvements in UC4 KPIs are expected.

Although the collection of the inspection documents by the security officer is still in progress, the initial tests have shown a noticeable improvement in the KPIs with TALON. For example, initially it takes about 20-25 minutes for a human to monitor the 4 different areas, whereas with the drone it takes 4-6 minutes. Furthermore, while the human is monitoring the sites for example every 3 hours, unlike the drone can do it every hour or even more often. This leads to a significant reduction in the time it takes to raise an alarm. For example, if the security officer conducts his rounds at 9:00 and then at 12:00 in the morning and an incident occurs at 11:00, the alarm will be raised after 12:00, when he starts his inspection. Unlike the drone, it can be done an hour earlier because it is able to conduct more rounds in the same timeframe.

Alongside the above KPI process in order to support the scenario, the company allowed us to film during its operation to build a dataset that will be used by another project partner to train a deep learning model that will identify people who wear or not PPE in the context of UC4. This data collection process provides valuable insights about the use of PPE in daily operations and helps to create more reliable and efficient prediction models, as it is trained on images from areas similar to the real conditions that the TALON system will run in UC4. It will contribute to KPI_16 concerning the development of the computer vision model.

Hence, for the needs of creating the data set, after gathering all the people directly involved and having a diagram of the whole area, several scenarios were first constructed and then filmed using the drone. To ensure diversity in the dataset, which is necessary for optimal model performance, different drone paths were designed, the video was shoot under different angles, heights and lighting conditions. Furthermore, the total number of workers shown in the video was varied, the number of those wearing or not wearing any protective equipment, the colour of the equipment (e.g. blue or yellow helmet) and finally the type of equipment (e.g. wearing extra vests). At the same time, the process was recorded by an independent person. The collected data is currently being processed and annotated and once completed will be used by the respective technical partners to train the model. Below is an overview of the process of annotating the images that were shoot (see *Figure 11*):

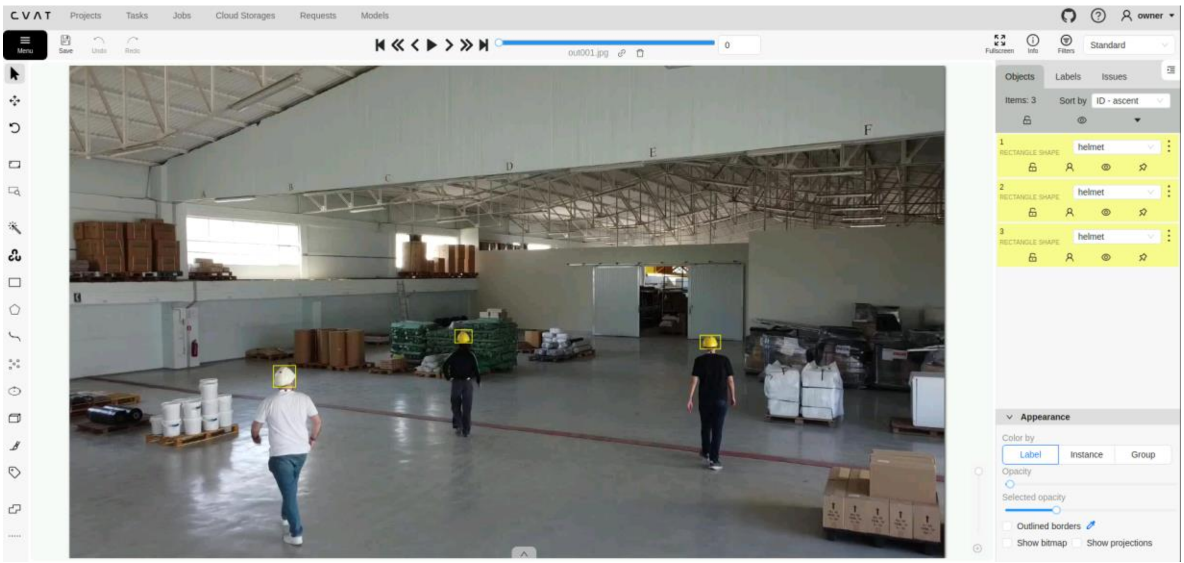

Figure 11: Annotation process

4 Conclusion and Future Outlook

Deliverable D5.3 – Pilot Specific TALON Platform Setup, Operation, Continuous Integration & Maintenance Report documents the work carried out in the TALON Pilot sites, undertaken under Task 5.3: Automatic UATV Coordination, Task 5.4: I5.0 Automation & Planning, Task 5.5: AR/VR for Training & Maintenance and Task 5.6: Human Robot Collaboration.

The current document presented the current status of deployment in each demonstrator together with a roadmap towards the successful delivery of the TALON Integrated platform. Additionally, a first evaluation of the TALON results has been conducted through an early evaluation of the KPIs defined for each use case. An updated report of these evaluation results will be delivered through D5.4 together with a complete evaluation of the technical results of the project.

D5.3 is the third out of four deliverables related to the “Work Package 5: Integration, Validation & Demonstration” of TALON, which marks the successful completion of Milestone 6 (MS6), related to the availability of TALON demonstrators in M25. It will be followed by “D5.4 Final TALON Platform Setup, Operation, Continuous Integration & Maintenance Report” which will mark the final release of the TALON Integrated platform providing additional information on the validation and evidence how the KPIs have been achieved.

Funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101070181