EVALUATION AND LESSONS LEARNED REPORT
D6.1

Editor Name	Roland Sitar, Ciprian Alexandrescu (FLEX)
Submission Date	31st August, 2025
Version	1.0
State	Final
Confidentially Level	PU

Co-funded by the Horizon Europe
Framework Programme of the European Union

EXECUTIVE SUMMARY

This report reviews how the FAIRWork approach and the DAI-DSS technology prototype performed when tested in real work environments. It examines whether these tools can support important workplace goals such as trust in decision-making, fairness, transparency, and overall efficiency. By evaluating these key aspects, the report assesses not only how the prototypes fit into the specific use case sites but also their potential impact in other industries and applications.

In particular, the evaluation focuses on three use cases in manufacturing environments, each addressing different operational challenges. First, the "Improve Information Access to Support Maintenance" use case, led by FLEX Althofen, aims to enhance maintenance efficiency by using AI to provide real-time, step-by-step guidance to technicians when diagnosing and repairing machine malfunctions. This use case contributes to the vision of **"Using AI to enable a broken machine to tell the maintenance staff how to efficiently fix itself"**. Second, the "Support Validation of Calibration Documents" use case, led by FLEX Timisoara, applies AI to automate the verification of calibration certificates, ensuring machines meet compliance standards by extracting and validating critical data from PDF documents. It contributes to the vision of **"Using AI to enable a machine to state if the machine is compliant using certification of calibration documents"**. Third, the "Assist Decisions about Production Planning" use case, led by CRF, uses an AI to assist planners in optimizing production schedules by integrating multi-parameter constraints and operator preferences. This use case contributes to the vision of **"Using AI to enable human operators to influence a multi-parameter optimization production planning based on personal preferences"**.

Also, the **DAI-DSS approach** and the overall project is evaluated, reflecting earlier findings based on practical experience, feedback and lessons learned. It considers the methods, prototypes, system architecture, exploitation strategies and the Innovation Shop and iterations to achieve the current state. The goal is to ensure that the findings remain reliable and applicable, while highlighting necessary improvements for future development.

Finally, the report presents a **roadmap for the future** of the use case partners for the planned operational adoption of the DAI-DSS. This roadmap lays out the steps needed to evolve both FAIRWork and DAI-DSS into operational solutions that can be broadly adopted in various sectors. It aligns with the overall exploitation strategy of the project, ensuring that the insights and improvements identified will drive future developments and successful industry integration.

By combining evaluations with forward-looking roadmaps, this report serves as an input for researchers, industry practitioners, and decision-makers interested in AI-driven workplace solutions.

PROJECT CONTEXT

Work package	WP6: Evaluation of FAIRWork
Task	T6.1: Evaluation of FAIRWork Use Case Scenarios T6.2: Evaluation of DAI-DSS Approach T6.3: Lessons Learned and Long-Term Impact
Dependencies	WP2, WP3, WP4, WP5, WP8

Contributors and Reviewers

Contributors	Reviewers
Alessandro Cisi, Giulio Vivo (CRF)	Antiona Markus (RWTH)
Roland Sitar, Ciprian Alexandrescu (FLEX)	Damiano Falcioni (BOC)
Noushin Qeybi, Stefan Böschen (RWTH)	Herwig Zeiner (JR)
Marlene Mayr (BOC)	
Sylwia Olbrych, Johannes Zysk (RWTH)	
Herwig Zeiner, Lucas Paletta (JR)	
Christian Muck (OmiLab)	

Approved by: Robert Woitsch [BOC], as FAIRWork Coordinator

Version History

Version	Date	Authors	Sections Affected
1.0	31st August, 2025	Roland Sitar, Ciprian Alexandrescu	All

Copyright Statement – Restricted Content

This document does not represent the opinion of the European Community, and the European Community is not responsible for any use that might be made of its content.

This is a restricted deliverable that is provided to the community under the license Attribution-No Derivative Works 3.0 reported defined by creative commons http://creativecommons.org

You are free:

	to share within the restricted community — to copy, distribute and transmit the work within the restricted community
Under the following conditions:	
	Attribution — You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).
	No Derivative Works — You may not alter, transform, or build upon this work.

With the understanding that:

Waiver — Any of the above conditions can be waived if you get permission from the copyright holder.

Other Rights — In no way are any of the following rights affected by the license:

- o Your fair dealing or fair use rights.
- o The author's moral rights.
- o Rights other persons may have either in the work itself or in how the work is used, such as publicity or privacy rights.

Notice — For any reuse or distribution, you must make clear to others the license terms of this work.
This is a human-readable summary of the Legal Code available online at:

http://creativecommons.org/licenses/by-nd/3.0/

TABLE OF CONTENT

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Meaning
AI	Artificial Intelligence
ANN/ NN	Artificial Neural Networks
CC	Calibration Certificate
CDB	Calibration Data Based
CP	Constraint Programming
DAI-DSS	Democratic Artificial Intelligence – Decision Support System
DoA	Description of Action
GUI	Graphical User Interface
HMI	Human Machine Interface
HTTP	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secure
HW	Hardware
KER	Key Exploitable Result
LLM	Large Language Model
MAS	Multi-Agent System
MCTS	Monte Carlo Tree Search
ML	Machine Learning
RL	Reinforced learning
RAG	Retrieval Augmented Generation
SSH	Secure Shell
SSL	Secure Sockets Layer
SW	Software
UI	User Interface
URS	User Requirement Specification

1 INTRODUCTION

1.1 Purpose of the Document

This document presents all validation procedures carried out using the FAIRWork DAI-DSS prototype. The primary objective of these activities is to assess the system's usability, operational effectiveness, and its overall impact on the AI-assisted support process.

FAIRWork aims to introduce AI-based decision support tools in industrial environments using the DAI-DSS (Democratized AI Decision Support System), which enables intelligent assistance in complex operational contexts. FAIRWork's DAI-DSS High-Level Architecture consists of five main components a) configurator, b) orchestrator, c) user interface, d) knowledgebase and e) AI-enrichment. The OLIVE platform as one element of the architecture is used to configure the Democratic AI Decision Support System (DAI-DSS) prototype for the relevant use case allowing diverse user groups to engage with and benefit from intelligent support processes in industrial and organizational contexts. This modular approach enables scalable and flexible deployment of AI-driven decision support functionalities.

Objectives

- To validate the usability and effectiveness of the DAI-DSS prototype.
- To assess the impact of AI-supported decision processes on operational efficiency.
- To ensure the system's adaptability across various industrial use cases.

Expected Outcomes

- A robust, user-friendly AI decision support system.
- Increased transparency and accessibility of AI tools in industrial settings.
- A scalable framework for future AI integrations.

In parallel to the use case evaluation of the DAI-DSS, the project and DAI-DSS approach is evaluated in particular by exploring the complex socio-technical challenges of implementing democratic decision-making (doing democracy) within companies and to understand the requirements for deploying a Decision Support System (DSS), a case study approach was used. The lead case was FLEX, located in Althofen, Austria. The study followed a three-phase process:

- Onboarding (October 2023): Initial site visit to FLEX for data collection through document analysis, observation, and employee interviews.
- Precision (July 2024): A deeper investigation at FLEX to understand human factors and democratization dynamics, including further interviews and a workshop.
- Contextualization (Fall 2024 & Summer 2025): Validation of findings through workshops and interviews to assess their applicability and feasibility for industry and service partners.

Also, reflection and lessons learned from the FAIRWork project is provided It reflects on challenges faced, effective strategies used, and adjustments made during the project. Additionally, the use case partners present their roadmaps for future development.

1.2 Document Structure

This report provides a structured and detailed account of the DAI-DSS demonstration conducted within the scope of the FAIRWork research project. It aims to document the methodological approach, evaluation procedures, and empirical findings, while situating the work within the broader scientific and industrial context of AI-supported decision-making in production environments.

The document is organized as follows:

- Chapter 1 introduces the objectives of the report and contextualizes the demonstration within the FAIRWork project framework.
- Chapter 2 presents the use case, including the applied evaluation methodologies and the corresponding results, with a focus on empirical validation.
- Chapter 3 offers a critical assessment of the DAI-DSS approach developed by RWTH Aachen, including methodological considerations and performance metrics, lessons learned and reflections on the DAI-DSS
- Chapter 4 discusses the anticipated long-term implications and future roadmap for industrial deployment.

The concluding section summarizes the key findings, outlines the benefits and limitations identified throughout the demonstration, and formulates recommendations for future implementation and scalability of the DAI-DSS system in real-world production settings.

References to relevant literature and project documentation are provided at the end of the report.

To ensure methodological transparency and comparability, the case studies of CRF and FLEX are presented in parallel. Accordingly, several sections are subdivided into dedicated subsections, each addressing the respective use case in a consistent and structured manner.

2 EVALUATION OF FAIRWORK USE CASE SCENARIOS

To ensure a comprehensive assessment of the DAI-DSS prototype for the use cases 1) Improve Information Access to Support Maintenance, 2) Support Validation of Calibration Documents and 3) Assist Decisions about Production Planning, a mixed-methods evaluation approach is applied. This methodology combines both quantitative and qualitative techniques to capture user feedback, operational impact, and overall effectiveness. In particular, the use case partners applied the following methods:

Improve Information Access to Support Maintenance (FLEX Althofen)

- Interviews
- Test Runs
- On-on one discussions with maintenance personals

Support Validation of Calibration Documents (FLEX Timisoara)

- On-on one discussions with calibration experts
- Test Runs

Assist Decisions about Production Planning (CRF)

- Questionnaires
- Interviews with operators
- Test Runs

Together, these methods provide an evaluation framework, ensuring that both technical performance and human factors are considered in the validation of each use case. In addition, a cost-benefit analysis was prepared to compare the expected benefits — such as increased productivity, reduced downtime, and improved resource utilization—against the associated costs, including implementation, training, and maintenance expected by each partner.

2.1 Improve Information Access to Support Maintenance (FLEX Althofen)

2.1.1 Use Case Description

"Use AI to enable a broken machine to tell the maintenance staff how to efficiently fix itself."

Abstract

In highly automated production settings, the complexity and variety of machinery present major challenges to maintaining efficiency and ensuring operational reliability. At Flex Althofen, an Electronic Manufacturing Services (EMS) provider, around 8,000 unplanned repair interventions occur annually—primarily due to increasing machine vulnerability and the limited experience of newly assigned staff. This paper introduces an AI-driven support system aimed at empowering machine operators, easing the burden on maintenance teams, and boosting Overall Equipment Effectiveness (OEE). By consolidating existing knowledge sources—such as maintenance logs, collaborative wikis, and equipment manuals—into a smart, unified interface, the system delivers context-sensitive guidance with minimal training effort. The solution promotes accessible knowledge and operational resilience, offering a scalable model for similar industrial environments.

Introduction

The increasing complexity of automated production systems in the electronics manufacturing sector has led to a rise in maintenance demands and machine breakdowns. At Flex Althofen, the annual volume of unplanned repair interventions has reached approximately 8,000 cases, significantly impacting production continuity and resource allocation. This situation is further aggravated by the growing number of inexperienced operators, whose limited technical knowledge often results in extended downtimes and frequent reliance on the maintenance team for minor issues.

To address these challenges, there is a clear need for a support system that empowers machine operators to independently resolve routine problems, thereby reducing the burden on maintenance personnel and improving overall system availability. Artificial Intelligence (AI) offers promising capabilities in this context, particularly in the form of intelligent decision support systems (DSS) that can provide real-time, context-sensitive guidance based on existing knowledge repositories.

The proposed approach involves the development of an AI-powered support tool that consolidates and intelligently interprets maintenance-related information from multiple existing sources:

- **ISPRO-NG Maintenance Database**: Serves as the central task management system, containing structured data such as maintenance instructions, task durations, and intervention logs.

- **Maintenance Wiki**: A collaborative platform used by engineers to document recurring issues and their respective solutions.

- **Equipment Documentation Repository**: Includes manuals, technical specifications, and operational guidelines for all machinery.

The AI system will be designed to:

1. **Aggregate and harmonize data** from the above sources using natural language processing (NLP) and semantic search techniques.

2. **Enable intuitive user interaction** through a conversational interface that requires minimal training and supports multilingual input.

3. **Provide context-aware recommendations** for troubleshooting and maintenance tasks, tailored to the specific machine, error type, and user profile.

4. **Continuous learning and improvement** through feedback loops and usage analytics, enhancing its accuracy and relevance over time.

The implementation will follow modular architecture to ensure scalability and integration with existing IT infrastructure. A pilot phase will be conducted at Flex Althofen, focusing on high-frequency breakdown scenarios and evaluating the system's impact on downtime reduction, operator autonomy, and maintenance team efficiency.

2.1.2 Use Case Evaluation

First, the knowledge base was created, and initial test were performed. Relevant documents were added to the knowledge base to create a solid foundation for AI-assisted decision support.

Initial experiments with the DAI-DSS tool were conducted, testing various questions to evaluate functionality and response rates. Discussions were held with the global IT team to enable the tool's activation on the corporate network and ensure broad access.

Second, the results from Step 1 were presented to the project partners and discussed. Based on the feedback received, the AI tool will be further developed to further improve its performance and practical usability. For an overview Figure 1 summarizes the evaluation approach with respect to the first and second prototypes' ability to provide accurate responses was evaluated taken a three-level based approach:

Level of System Evaluation	Objective	Stage of Prototype	Output accuracy
Level 1 Evaluation with Standardized Questions	Test the ability to deliver consistent responses to identical questions across different input modalities	First prototype	100%
Level 2 Evaluation with Topic-Based Free Questions	Test the ability to understand and respond accurately to freely formulated questions within predefined topic areas	First prototype	68%
		Second prototype with extended database & additional user training	84%
Level 3 Evaluation with Open Problem Statements	Test the flexibility and problem-solving capabilities when AI faces open-ended, user-defined challenges.	First prototype	46%
		Second prototype with extended database & additional user training	68%

Figure 1 Overview of evaluation approach FLEX Althofen

2.1.2.1   Initial Evaluation and System Introduction (Phase 1)

During the first project phase, the initial prototype was successfully developed and deployed. The main objective was to gain a foundational understanding of the system's architecture and functionality through targeted testing and direct interaction.

Conducted Tests and Results:

System Initialization and Core Functionality:

The prototype booted without errors. All core functions were successfully initialized.

Platform Access Restriction:

One notable disadvantage encountered during testing was the requirement to unlock the Olive platform through BOC for each session. This dependency introduced delays and limited autonomous access, which may impact scalability and operational efficiency in future deployments.

User Interface and Usability:

The graphical user interface (GUI) proved to be intuitive and stable.

Conclusion and Outlook:

The results of the initial evaluation confirm the basic functionality and stability of the system. These insights provide a solid foundation for further development, particularly in optimizing fault detection algorithms, expanding sensor capabilities, and enhancing user interaction. The next phase will focus on refining system performance and preparing for integration into a broader operational environment.

2.1.2.2 Evaluation Process (Phase 2)

Following the development of the second prototype, a structured multi-level evaluation was conducted to assess the system's performance across different input formats and levels of complexity. The evaluation aimed to measure the accuracy, consistency, and adaptability of the AI tool in both written and spoken interactions.

Level 1: Evaluation with Standardized Questions

To test the system's ability to deliver consistent responses to identical questions across different input modalities.

Five participants each submitted ten identical questions using three formats:

- Written input.
- Spoken input in English.
- Spoken input in German.

The written questions were strictly identical to ensure a controlled testing environment.

Spoken inputs were recorded and analyzed qualitatively but not included in the quantitative scoring.

Target Outcome:

- Written input: 100% consistency in question formulation and response accuracy.
- Spoken input: Responses were documented for reference but not scored.

Level 2: Evaluation with Topic-Based Free Questions

To evaluate the system's ability to understand and respond accurately to freely formulated questions within predefined topic areas.

Each of the five participants asked 2 freely formulated questions on five predefined topics, using:

- Written input.
- Spoken input in English.
- Spoken input in German.

This level introduced variation in phrasing while maintaining thematic consistency.

Target Outcome:

- Written input: At least 80% of the responses should be accurate and relevant to the topic.
- Spoken input: Responses were recorded and reviewed qualitatively.

Level 3: Evaluation with Open Problem Statements

To test the system's flexibility and problem-solving capabilities when faced with open-ended, user-defined challenges.

Each participant submitted ten freely chosen problem statements, using:

- Written input.
- Spoken input in English.
- Spoken input in German.

This level allowed for maximum variability in content and complexity.

Target Outcome:

- Written input: At least 60% of the responses should be correct, coherent, and solution oriented.
- Spoken input: Responses were documented but not quantitatively assessed.

2.1.2.3 Technical Questions for Evaluation Purposes

✅ **Test 1–3: General Technical Questions**

• What can I do when PPM values are too high?
• The scanner is not working – what troubleshooting steps can I take?
• Why is the splice not accepted, and how can I fix it?
• The piezo capacity is too low – what could be the reason?
• The screen cleaner does not move up and cleaning cannot be performed – what can I do?
• How do I replace a filter disc correctly?
• How do I properly clean a pressure regulator valve?
• Why does pipette scanning fail, and how can I resolve it?
• The cartridge is difficult to remove from the holder unit – what could be the cause?
• How do I create a complete and clear bug report?

✅ **Test 4–6: Topic-Based Questions**

• Pick & Place Head Replacement:
• How do I safely and correctly replace a Pick & Place head?
• X-ray Inspection: What are common issues with the X-ray system and how can they be resolved?
• AOI (Automated Optical Inspection): What should I do if the AOI system misidentifies components or positions?
• Reflow Process: What actions should I take when problems occur during the reflow process?
• Paste Inspection: How can I detect and correct errors during solder paste inspection?

✅ **Test 7–9: Operator-Specific Questions**

👷 Operator 1:

▪ Which lubricant is used for the SMT oven transport chain?
▪ How do I generate a bug report from the placement machine?
▪ How can I unbind a PCB from the process?
▪ What should I do if the SMT oven transport chain goes into overload?
▪ Where can I find the user manual for the Mycronic MY700 dispensing machine?
▪ What should I do if the placement head temperature is too high?
▪ How often should maintenance be performed on the Intelijet?
▪ What is the minimum required laser power for an Osai laser cutter?
▪ How do I perform a dispense test on the Yamaha YSD?
▪ Which pressure regulator valve is used in a CPP placement head?

👷 Operator 2:

▪ How do I replace a CPP head correctly?
▪ How do I calibrate the D1 labeler?
▪ What can I do if the PCB sticks to the stencil after printing?
▪ How do I clean a DRV (pressure regulator valve)?
▪ How should the height measurement points be set on the MY system?
▪ The screen cleaner on the TQ does not move up – what can I do?
▪ How do I configure the barcode settings on the SPI system?
▪ What can I do about small glue dots on the Yamaha dispenser?
▪ What can I do about unjustified failures in the X-ray system?
▪ What must be considered when replacing a CPP head?

👷 Operator 3:

▪ What should I do in case of encoder counting errors?
▪ How do I handle a grabber error?
▪ What should I do if the ISS is not reachable?
▪ What actions are needed if the FCU becomes hot?
▪ What should I do in case of misaligned printing on the DEK?
▪ How do I resolve a vacuum error on the X4 system?
▪ What should I do if the X-ray system does not detect the PCB?
▪ What should I do if the AOI system misidentifies components?
▪ What does "bareboard teaching" mean and how is it performed?
▪ What should I do if solder paste residues remain on the PCB?

👷 Operator 4:

▪ What should I do in case of misaligned components?
▪ What should I do if the vacuum pump on the X4 is noisy?
▪ How do I resolve vacuum issues on the CPP head?
▪ What should I do if the pin picker is outside the travel range and cannot pick up components?
▪ What should I do if the SX2 cannot perform reference positioning?
▪ What should I do in case of encoder counting errors?
▪ What should I do if the machine discards components during placement?
▪ What should I do if the transport width cannot be adjusted on the X4s?
▪ What should I do if the light barrier does not detect the PCB during SX2 transport?
▪ What should I do if the stationary camera does not capture images?

🧑‍🏭 Operator 5:

▪ How do I correctly set the fiducial marks on the LMC?
▪ What can I do about high PPM values on the placement machine?
▪ What can I do if the SPI system fails due to excessive solder paste?
▪ What can I do about misaligned components from the placement machine?
▪ The KIC system on the oven shows red – what does it mean and what should I do?
▪ The SMT oven shows error "Node ID 31 Overload" – how do I resolve it?
▪ I cannot reach the target throughput – what can I do to improve it?
▪ What can I do about missing glue dots from the Mycronic dispenser?
▪ What should I do about recurring transport errors on the Mycronic system?
▪ I'm experiencing many process errors on the placement machine – what could be the cause?

2.1.2.4 Output of the Evaluation

The evaluation of the speech recognition module was conducted within the active production environment to ensure realistic operational conditions. However, the presence of continuous and variable background noise—originating from industrial machinery, human activity, and environmental acoustics—resulted in significant degradation of the audio signal-to-noise ratio.

This interference led to inconsistent recognition accuracy, increased false positives/negatives, and reduced reliability of voice-triggered commands. Due to the non-deterministic nature of the acoustic disturbances and the inability to isolate clean audio input, the speech module could not be reliably assessed under these conditions.

Consequently, the scope of this report is limited to the evaluation of system responses based on written input modalities (e.g., keyboard, touchscreen), which provided controlled and reproducible interaction data for analysis.

Level 1: Evaluation with Standardized Questions

In this round, all questions were predefined and had to be entered into the system according to a specific format. This ensured consistency in data input and allowed for a structured evaluation of the system's ability to process and respond to standardized queries. The use of fixed question parameters also helped to minimize variability and focus the assessment on system performance rather than user interpretation.

Test 1		Exact same question ask 10 question in written form								1=correct answer	
										0=wrong answer	
Operator		Question 1	Question 2	Question 3	Question 4	Question 5	Question 6	Question 7	Question 8	Question 9	Question 10
1		1	1	1	1	1	1	1	1	1	1
2		1	1	1	1	1	1	1	1	1	1
3		1	1	1	1	1	1	1	1	1	1
4		1	1	1	1	1	1	1	1	1	1
5		1	1	1	1	1	1	1	1	1	1
				Result	**100,00%**						

Figure 2 Output level 1

The test level in question was successfully completed with a 100% success rate. The results achieved proved consistent and reproducible in multiple independent runs, confirming the reliability and validity of the AI.

Level 2: Evaluation with Topic-Based Free Questions

Run1

Test 4		Ask 2 freely chosen questions on the same 5 topic in written form							1=correct answer	
									0=wrong answer	
		Topic 1		Topic 2		Topic 3		Topic 4		Topic 5	
Operator		Question 1	Question 2	Question 1	Question 2	Question 1	Question 2	Question 1	Question 2	Question 1	Question 2
1		0	1	1	0	1	1	1	1	1	1
2		1	1	1	0	1	0	0	0	0	0
3		0	0	1	1	1	1	1	0	1	1
4		0	0	1	1	1	1	1	1	1	1
5		1	1	1	0	1	1	1	0	0	1
				Result	68,00%						

Figure 3 Output level 2 -Run 1

The first run in this level fell short of expectations. As part of the evaluation, it was checked whether the relevant content was already available in the knowledge database. Missing information was identified and subsequently uploaded to fill the gaps.

To improve the quality of the results, the test participants were trained on how to manage topic-specific questions more effectively. This aimed to enhance their understanding and lead to more accurate and meaningful outcomes.

Run 2

Test 4		Ask 2 freely chosen questions on the same 5 topic in written form							1=correct answer	
									0=wrong answer	
		Topic 1		Topic 2		Topic 3		Topic 4		Topic 5	
Operator		Question 1	Question 2	Question 1	Question 2	Question 1	Question 2	Question 1	Question 2	Question 1	Question 2
1		0	1	1	0	1	1	1	1	1	1
2		1	1	1	0	1	1	0	1	1	1
3		1	1	1	1	1	1	1	0	1	1
4		0	1	1	1	1	1	1	1	1	1
5		1	1	1	0	1	1	1	0	1	1
				Result	84,00%						

Figure 4 Output level 2 -Run 2

The final test run conducted in July 2025 was included in the overall performance assessment of the initiative. The results not only met but exceeded the predefined target metrics, indicating a significant improvement compared to previous levels.

This positive outcome reflects the effectiveness of the adjustments made during earlier stages, including the enrichment of the knowledge database and the targeted training of test participants. The successful completion of this level marks an important milestone and provides a solid foundation for the next steps in the project.

Level 3: Evaluation with Open Problem Statements

The objective of this study is to evaluate the system's flexibility and problem-solving capabilities when confronted with open-ended, user-defined challenges.

Each participant submitted ten freely chosen problem statements, formulated in written form. These tasks served as the basis for assessing the system's responses in terms of creativity, solution quality, and adaptability to diverse requirements.

First Trial

Test 7		Ask freely chosen questions any topic in written form						1=correct answer			
								0=wrong answer			
Operator		Question 1	Question 2	Question 3	Question 4	Question 5		Question 6	Question 7	Question 8	Question 9	Question 10
1		1	0	0	1	1		1	0	0	1	0
2		0	0	1	0	1		0	0	1	0	1
3		0	0	1	1	1		1	1	1	0	0
4		1	0	0	0	1		1	1	1	1	0
5		0	0	0	0	0		1	0	0	1	0
				Result	46,00%						

Figure 5 First Trail in level 3

The results of the first evaluation round were significant below expectations. Upon analysis, two main reasons for this outcome were identified:

Deviation from the Defined Scope:

A few of the submitted problem statements did not align with the intended focus area, which was limited to SMT (Surface-Mount Technology) production. This misalignment led to irrelevant or unproductive system responses, as the challenges fell outside the domain for which the system was designed and trained.

Insufficient Knowledge Base:

Like the issues observed in level 2, the system's knowledge database lacked the necessary depth and specificity to adequately address the submitted tasks. The limited coverage of SMT-related content restricted the system's ability to generate accurate, insightful, and contextually appropriate solutions.

Improved Evaluation Results (Run 5):

Following several iterations and targeted adjustments, the system's performance improved significantly, with results consistently approaching or exceeding the defined target benchmarks. The most successful outcomes were achieved using the set of questions outlined in Point 2.1.2.3, which were specifically tailored to the SMT production domain.

These questions demonstrated a clear alignment with the system's capabilities and the refined scope of the evaluation. The improvements can be attributed to both the narrowing of the focus to relevant topics and the gradual enhancement of the underlying knowledge base. Together, these factors enabled the system to deliver more accurate, context-aware, and actionable responses.

This level of the evaluation confirms that with well-defined input and a sufficiently developed knowledge infrastructure, the system is capable of meeting exacting standards in problem-solving and domain-specific support.

Test 7		Ask freely chosen questions any topic in written form					1=correct answer			
							0=wrong answer			
Operator		Question 1	Question 2	Question 3	Question 4	Question 5	Question 6	Question 7	Question 8	Question 9	Question 10
1		1	1	1	1	1	1	0	0	1	0
2		1	0	1	1	1	0	0	1	0	1
3		0	1	1	1	1	1	1	1	0	0
4		1	0	1	0	1	1	1	1	1	0
5		1	1	0	1	0	1	0	1	1	1
					Result	68,00%					

Figure 6 Best result in level 3

Response time of the system

Throughout all test iterations, the measured response time remained below the threshold of 10 seconds, thereby fulfilling the requirements for acceptable system performance as defined by the evaluation

2.1.3   Cost-Benefit Analysis

At this stage, we can only present an estimated simulation of potential cost savings. This is primarily since the observation period has been too short to draw reliable conclusions. Additionally, the phase of reduced production output was also limited in duration, which further restricts the significance of the available data.

Another key factor is that the AI tool has not yet been rolled out to the operator level. As a result, potential effects on operator autonomy and the reduction of unplanned maintenance interventions have not yet been fully captured or evaluated.

- Reduction of tickets, cost and time saving in the maintenance department
- System Load Due to Unplanned Tickets

Currently, the system is managing approximately 8,000 unplanned service tickets, which have been automatically logged due to various machine-related incidents. These tickets typically originate from unexpected breakdowns, operator interventions, or unresolved maintenance issues. The high volume indicates a significant strain on support resources and highlights the need for improved predictive maintenance and operator-level diagnostics.

Following a detailed evaluation of the logged service orders, it was determined that approximately 30% of the unplanned tickets were caused by minor issues or operator errors. These types of incidents typically do not require intervention from maintenance personnel and could be resolved directly by the machine operators.

With the support of an AI-based assistance tool, such cases could be managed autonomously by the operators, eliminating the need to generate a service ticket. This would not only reduce system load but also improve response times and enhance operator competence in handling routine disruptions.

Extrapolated over a full year, this corresponds to approximately 2,400 service tickets that do not require intervention by the maintenance team. These tickets could be resolved directly by the operators themselves, provided they are supported by an AI-based assistance system.

Implementing such a solution would significantly reduce the workload on maintenance personnel, improve system efficiency, and empower operators to manage minor disruptions independently. This not only streamlines operations but also contributes to faster issue resolution and reduced machine downtime.

Cost Calculation – Maintenance Manpower & Line Downtime

1. Maintenance Manpower Savings

- Ticket Reduction: 2,400 tickets/year
- Effort per Ticket: 0.75 hours (including walking time to machine)
- Hourly Rate: €25/hour (maintenance personnel)

- **Annual Savings**
  $2,400 \times 0.75\,h \times €25/h = €45,000$
  =€45,000€

Faster Response to High-Priority Issues
Reduced backlog and quicker reaction to urgent breakdowns.
Improved uptime for key production lines.

2. Line Downtime Savings

- Downtime Reduction per Case: from 0.75h to 0.60h → Δ = 0.15h
- Downtime Cost: €250/hour

- **Annual Savings**
  $2,400 \times 0.15\,h \times €250/h = €90,000$
  =€90,000€

Energy Consumption
Unplanned Downtime = Energy Waste
When machines stop unexpectedly and need to be restarted, this often leads to inefficient energy cycles. For example: Motors running idle.
Machines may continue to consume power even though no productive work is being done.
Heating or cooling processes need to be restarted.
Thermal systems require significant energy to return to operating temperature, especially if they were interrupted mid-cycle.
Conveyor systems consume energy without transporting material.
Conveyors and other auxiliary systems may continue operating unnecessarily, leading to wasted energy.
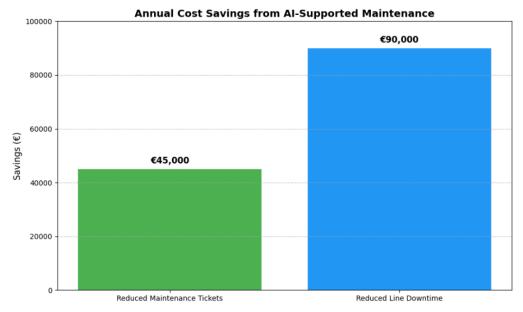
Figure 7 Cost saving overview

The presented cost savings are estimates based on preliminary data.

Due to the limited observation period, no statistically validated or operationally confirmed values could be collected.

Further analysis over an extended timeframe is required to derive reliable and representative figures.

As indirect effect, using the AI prototype improved the understanding of AI technologies in FLEX Althofen and enabled employees to deepen their knowledge for using AI.

2.1.4 Summary of Findings

The speech recognition module was assessed in a live production environment. However, due to constant and unpredictable background noise from machinery and human activity, the audio signal quality was significantly degraded. As a result, voice commands were unreliable, and the evaluation focused instead on written input methods (e.g., keyboard, touchscreen), which provided consistent and reproducible data.

Level 1: Standardized Questions

Predefined questions were entered in a fixed format.

Result: 100% success rate. Responses were consistent and validated across multiple test runs.

Level 2: Topic-Based Free Questions

Run 1: Performance was below expectations due to missing knowledge base content. The gaps were filled, and participants received training.

Run 2: Marked improvement. Results exceeded target metrics thanks to database enrichment and participant training.

Level 3: Open Problem Statements

Objective: Assess the system's flexibility and problem-solving ability with user-defined tasks.

Method: Each participant submitted ten freely formulated written challenges.

First Trial: Results were unsatisfactory due to:

- Tasks falling outside the intended SMT (Surface-Mount Technology) domain.
- Insufficient depth in the knowledge base.

Run 5: After several iterations and refinements, performance improved significantly. The best results came from questions aligned with the SMT domain and the system's capabilities.

Based on the results obtained throughout the evaluation levels, it is evident that the tool demonstrates strong potential for future integration into operational workflows. Its ability to deliver dependable and context-aware responses—particularly when supported by a well-structured and domain-specific knowledge base—highlights its value as a digital assistant in industrial environments.

To fully realize this potential, it is essential that the underlying knowledge database be systematically expanded to encompass all relevant machine areas. Only then can the system provide comprehensive support across diverse production domains and contribute meaningfully to process optimization, troubleshooting, and knowledge transfer.

With continued development and targeted enhancements, this tool could become a significant asset in driving efficiency and innovation within the organization.

Despite the promising aspects of the tool, one significant limitation was observed during the evaluation: the speech recognition module proved unsuitable for use within our production environment. The presence of continuous and unpredictable background noise—generated by industrial machinery and human activity—severely impacted the system's ability to accurately process voice commands.

As a result, the module failed to deliver reliable performance under real-world conditions, leading to inconsistent recognition outcomes and reduced usability. This limitation underscores the need for either substantial improvements in noise filtering and signal processing or a shift in focus toward more robust input modalities such as written commands.

Suggested Follow-Up Actions by the AI Tool

Knowledge Gaps Detection

Automatically identify missing content in the knowledge base based on user queries. Suggest relevant documents or data sources for upload.

Contextual Question Proposals

After answering a query, the system could propose related questions:

"Would you like to explore optimization strategies for this machine?"

"Do you want to compare this result with similar equipment?"

Interactive Troubleshooting Paths

If a problem is detected, the system could guide the user through a step-by-step diagnostic process.

Example: "Based on your input, would you like to run a fault analysis on the XYZ module?"

Feedback Loop Integration

After each interaction, the system could ask:

"Was this answer helpful?"

"Would you like to refine your question?"

Cross-System Linking

Propose actions in connected systems (e.g., maintenance scheduling, ERP updates).

"Would you like to create a maintenance ticket for this issue?"

2.2 Support Validation of Calibration Documents (FLEX Timisoara)

2.2.1 Use Case Description

"Use AI to enable a machine to state if the machine is compliant using certification of calibration documents."

Abstract

Verification of Calibration Certificates (CC) in PDF format is currently a manual, repetitive, time-consuming and error-prone process. Each calibration certificate received from the calibration service provider must be carefully checked for any discrepancies or missing or incorrect information. Employees must manually check instrument details (model, type, manufacturer, serial number), calibration dates (calibration and expiry), calibration results (measured values, deviations, uncertainties), measuring equipment information (for traceability), and formal aspects (signatures, page numbering, completeness). The current manual verification of calibration certificates is a time-consuming, repetitive, and error-prone process that demands meticulous attention to critical data points his exhaustive manual review, exacerbated by the sheer volume of certificates, significantly increases the risk of human error. Overlooked discrepancies can have severe repercussions, potentially leading to the use of faulty measuring equipment and compromising product or service quality.

Introduction

To address these challenges and optimize the process, an AI-powered application was developed to automate the verification of Calibration Certificates. This application leverages AI to analyze certificates, identify discrepancies, and automatically generate a comprehensive report highlighting any missing or incorrect information. This automated solution is expected to drastically reduce review time and eliminate the likelihood of human error, leading to a much higher level of accuracy and reliability in the certificate review process. This will bolster internal quality assurance and allow staff previously dedicated to manual verification to be reassigned to more value-adding activities.

In Flex Timisoara five category of Calibration Certificate where identified. Four of them are structured following a specific template. The last category "Others" is unstructured and is out of scope for this project:
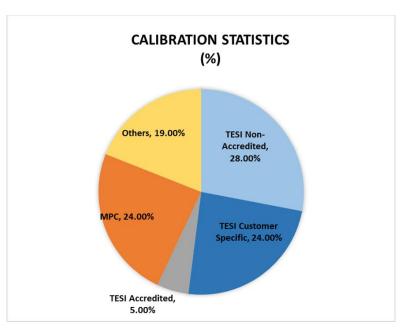
Figure 8 Category of Calibration Certificate

For these 4 templates (structured category) under study several 14 URSs (User Requirements Specifications) were defined:

URS	Name	Verification
URS01	Issue Date	not earlier then Calibration Date
URS02	Device Name	same as in CDB
URS03	Device Manufacturer	same as in CDB
URS04	Device Model	same as in CDB
URS05	Device Serial Number	same as in CDB
URS06	Device ID	same as in CDB
URS07	Calibration Date	same as in CDB
URS08	Calibration Due Date	same as in CDB
URS09	Standards Due Date	not later than Calibration Date
URS10	No. of Pages	presence of all pages
URS11	Overall Results	all summary results are Pass
URS12	Measurement Results	all individual measurements results are Pass
URS13	Executor Signature	presence of signature
URS14	Approver Signature	presence of signature

Figure 9 URS definition

In the case of the Calibration Certification Service, the application is a standalone application, a simply executable file running on a standalone PC allowing multiple file (docs) processing.

The application identifies and interprets data from the Calibration Certificate PDF file, localizing each field associated with the respective URS. It then compares these fields against our Calibration Database (CDB), which is represented as an Excel file reflecting our local SQL Database, to verify the presence and accuracy of information as specified for each URS in the table above.

A configuration file 'config.yml' is used to change between Calibration Certificate category type.

The Calibration Certification Service generates two types of reports:

1. **General one containing the summarization of the results of all certificates processed**

12/06/2025 17:20:39
Certificates tested in total: 60
    passed: 49 (81.67%)
    failed: 11

URS Failed in total:
    URS05: 1
    URS06: 8
    URS10: 1
    URS12: 2

URS Explanation/Expectations

URS01: Issue Date
URS02: Device Name
URS03: Device Manufacturer
URS04: Device Model
URS05: Device Serial Number
URS06: Device ID
URS07: Calibration Date
URS08: Calibration Due Date
URS09: Standards Due Date
URS10: No. of Pages
URS11: Overall Results
URS12: Measurement Results
URS13: Executor Signature
URS14: Approvar Signature

Overview of All Flagged Certificates

0277_2024_RO_E_thickness gauge.pdf

    Failed URSs:
      URS06: Device ID
    File Overview:

      Page 1
      [URS06] Failed.
        Calibration device not found. ID may be wrong.
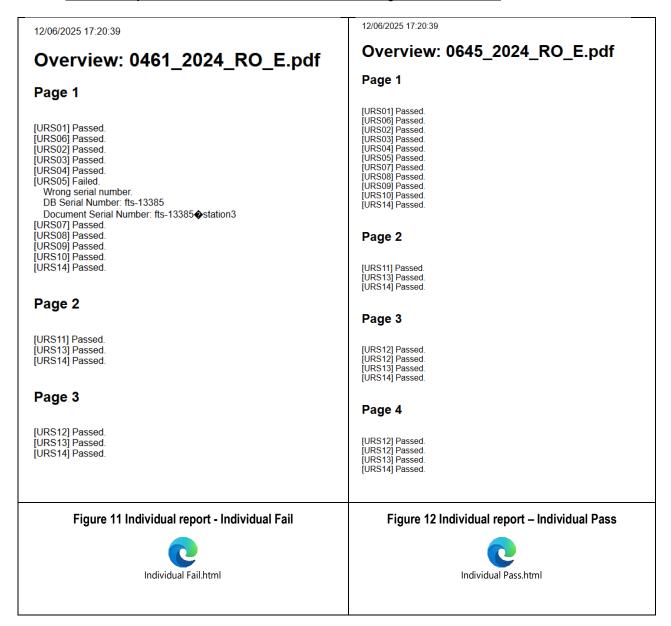
0461_2024_RO_E.pdf

    Failed URSs:
      URS05: Device Serial Number
    File Overview:

Figure 10 General Report

General.html

2. **Individual reports for all Calibration Certificate containing individual results.**

12/06/2025 17:20:39
 	
## Overview: 0461_2024_RO_E.pdf 	
 	
### Page 1 	
 	
[URS01] Passed.
[URS06] Passed.
[URS02] Passed.
[URS03] Passed.
[URS04] Passed.
[URS05] Failed.
   Wrong serial number.
   DB Serial Number: fts-13385
   Document Serial Number: fts-13385�station3
[URS07] Passed.
[URS08] Passed.
[URS09] Passed.
[URS10] Passed.
[URS14] Passed.
 	
### Page 2 	
 	
[URS11] Passed.
[URS13] Passed.
[URS14] Passed.
 	
### Page 3 	
 	
[URS12] Passed.
[URS13] Passed.
[URS14] Passed.
	12/06/2025 17:20:39
	
	## Overview: 0645_2024_RO_E.pdf
	
	### Page 1
	
	[URS01] Passed.
	[URS06] Passed.
	[URS02] Passed.
	[URS03] Passed.
	[URS04] Passed.
	[URS05] Passed.
	[URS07] Passed.
	[URS08] Passed.
	[URS09] Passed.
	[URS10] Passed.
	[URS14] Passed.
	
	### Page 2
	
	[URS11] Passed.
	[URS13] Passed.
	[URS14] Passed.
	
	### Page 3
	
	[URS12] Passed.
	[URS12] Passed.
	[URS13] Passed.
	[URS14] Passed.
	
	### Page 4
	
	[URS12] Passed.
	[URS12] Passed.
	[URS13] Passed.
	[URS14] Passed.

Figure 11 Individual report - Individual Fail Individual Fail.html	**Figure 12 Individual report – Individual Pass** Individual Pass.html

A configuration file .html is used to change between Calibration Certificate categories type.

2.2.2   Use Case Evaluation

In this chapter is described the methodology used in Flex Timisoara to evaluate the prototype and next revisions of Calibration Certificate Service/Application. This was an iterative process of improvement and validation of new versions released. We can identify 4 major steps:

Step 1

Running a small sample size (few randomly chosen Calibration Certificate) just to see the behavior (output) of the Prototype in terms of stability and repeatably.

Step 2

Running a large sample size (all batch of one category for year 2024), focusing on potential "falls calls" and other issues that appear when using a large sample size.

The issues faced during tests were summarized and track with CoC Action Tracker.

Observation: Evaluation is done for CoC from year 2024 and CDB used is an excel image of real SQL Calibration Data Base (CDB).

Step 3

To evaluate the stability and repeatability of the ST tool's output, the Attribute Study - MSA (Measurement System Analysis) technique was employed. Although this method is not typically utilized for assessing software capabilities, it was deemed suitable in this context, as the software tool is responsible for verifying and inspecting Calibration Certificates.

Technique used: Attribute Study

Goal: assessment of (inspection) Effectiveness, Probability of False Alarm, Probability of Miss

Sample size:

- 230 inspections of "good" samples/certificates
- 230 inspections of "non conform" samples/certificates.

Above sample size ensures a 90% / 99% Confidence / Reliability levels, considered suitable for the application.

Acceptance Criteria:

- Effectiveness: >= 0.9
- Probability of False Alarm: <=0.05
- Probability of Miss: 0

Calculation:

- Effectiveness: (N+X)/Ti
- Probability of False Alarm: Fn/(N+Fn)
- Probability of Miss: M/(X+M)

Where:

N = The total number of "good" samples correctly identified (accepted)

X = The total number of "non conform" samples correctly identified or rejected (rejected)

M = The total number of "non conform" samples accepted or classified as good. <Misses>

Fn = The total number of "good" samples rejected or classified as bad. <False Alarms>

Ti = Total number of inspections

In our cases to ensure the Confidence/ Reliability levels of 90% / 99% the following was used:

- Good / Bad sample ratio = 50%
- Number of bad samples (with known defects to cover all URSs) = 10
- Number of good samples = 10

- Number of appraisers = 1
- Number of repeats = 23

This means:

- Total no. inspections of known defects = 230 (minim: 229)
- Total no. inspections of known good parts = 230 (minim: 229)

"Good" samples are considered Calibration Certificates, previously verified by calibration technician to ensure no errors are present; expected output: SW Tool shall report that all certificates are ok, with no issues.

"Not good" samples are considered Calibration Certificates altered by generating known errors to challenge all 14 URSs; expected output: SW Tool shall report all errors.

Observation: appraisers = 1 because the result of running tool is not influenced by the person who made the test/study

The results of running the study were good as expected proving the repeatability stability of SW tool and are shown in excel below as well as in Figure 13 and Figure 14:

Cal ID	Certificate number	Equipment type	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Run 7
2911	0860_2024_RO_E	Power supply	PASS	PASS	PASS	PASS	PASS	PASS	PASS
2450	0864_2024_RO_E	Multimeter	PASS	PASS	PASS	PASS	PASS	PASS	PASS
10509	2599_2024_RO_E	Multifunction I/O Device	PASS	PASS	PASS	PASS	PASS	PASS	PASS
12252	0247_2024_RO_E	Electric Screwdriver	PASS	PASS	PASS	PASS	PASS	PASS	PASS
12830	0463_2024_RO_E	Load Cell	PASS	PASS	PASS	PASS	PASS	PASS	PASS
7889	1329_2024_RO_E	Scale	PASS	PASS	PASS	PASS	PASS	PASS	PASS
6334	2608_2024_RO_E	DC power supply	PASS	PASS	PASS	PASS	PASS	PASS	PASS
8549	0762_2024_RO_E	Universal Frequency Counter	PASS	PASS	PASS	PASS	PASS	PASS	PASS
6928	2825_2024_RO_E	Climatic Chamber	PASS	PASS	PASS	PASS	PASS	PASS	PASS
10578	0760_2024_RO_E	Caliper	PASS	PASS	PASS	PASS	PASS	PASS	PASS

Figure 13 Stability Study: Good Documents

Cal ID	Certificate number	Equipment type	URS modified	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6
2911	0860_2024_RO_E	Power supply	URS 01 /URS 02	URS Explanation	URS Explanation	URS Explanation	URS Explanation	URS Explanation	URS Explanation
2450	0864_2024_RO_E	Multimeter	URS 03 /URS 04						
10509	2599_2024_RO_E	Multifunction I/O Device	URS 05	URS01: Cert	URS01: Cert	URS01: Cert	URS01: Cert	URS01: Cert	URS01: Ce
12252	0247_2024_RO_E	Electric Screwdriver	URS 06	URS02: Calit	URS02: Calit	URS02: Calit	URS02: Calit	URS02: Calit	URS02: Ca
12830	0463_2024_RO_E	Load Cell	URS 07 / URS 08	URS03: Man	URS03: Man	URS03: Man	URS03: Man	URS03: Man	URS03: Ma
7889	1329_2024_RO_E	Scale	URS 09 / URS 10	URS04: Calit	URS04: Calit	URS04: Calit	URS04: Calit	URS04: Calit	URS04: Ca
8549	0762_2024_RO_E	Universal Frequency Counter	URS 11	URS05: Calit	URS05: Calit	URS05: Calit	URS05: Calit	URS05: Calit	URS05: Ca
10578	0760_2024_RO_E	Caliper	URS 12	URS06: Calit	URS06: Calit	URS06: Calit	URS06: Calit	URS06: Calit	URS06: Ca
6928	2825_2024_RO_E	Climatic Chamber	URS 13	URS07: Cert	URS07: Cert	URS07: Cert	URS07: Cert	URS07: Cert	URS07: Ce
6334	2608_2024_RO_E	DC power supply	URS 14	URS08: Cert	URS08: Cert	URS08: Cert	URS08: Cert	URS08: Cert	URS08: Ce
				URS09: Mea	URS09: Mea	URS09: Mea	URS09: Mea	URS09: Mea	URS09: Me
				URS10: Doc	URS10: Doc	URS10: Doc	URS10: Doc	URS10: Doc	URS10: Do
				URS11: All to	URS11: All to	URS11: All to	URS11: All to	URS11: All to	URS11: All
				URS12: All to	URS12: All to	URS12: All to	URS12: All to	URS12: All to	URS12: All
				URS13: Engi	URS13: Engi	URS13: Engi	URS13: Engi	URS13: Engi	URS13: En
				URS14: Hea	URS14: Hea	URS14: Hea	URS14: Hea	URS14: Hea	URS14: He
				Overview of All Flagged Certificates	Overview of All Flagged Certificates	Overview of All Flagged Certificates	Overview of All Flagged Certificates	Overview of All Flagged Certificates	Overview of All Flagged Certificates
				0247_2024_RO_E.pdf	0247_2024_RO_E.pdf	0247_2024_RO_E.pdf	0247_2024_RO_E.pdf	0247_2024_RO_E.pdf	0247_2024_RO_E.pdf
				Failed URSs:	Failed URSs:	Failed URSs:	Failed URSs:	Failed URSs:	Failed URSs:
				URS06: Calibration ID is valid.	URS06: Calibration ID is valid.	URS06: Calibration ID is valid.	URS06: Calibration ID is valid.	URS06: Calibration ID is valid.	URS06: Calibration ID is valid.
				File Overview:	File Overview:	File Overview:	File Overview:	File Overview:	File Overview:
				Page 1	Page 1	Page 1	Page 1	Page 1	Page 1
				[URS01] Pas	[URS01] Pas	[URS01] Pas	[URS01] Pas	[URS01] Pas	[URS01] Pa
				[URS06] Fail	[URS06] Fail	[URS06] Fail	[URS06] Fail	[URS06] Fail	[URS06] Fa
				Calibration d	Calibration d	Calibration d	Calibration d	Calibration d	Calibration
				0463_2024_RO_E.pdf	0463_2024_RO_E.pdf	0463_2024_RO_E.pdf	0463_2024_RO_E.pdf	0463_2024_RO_E.pdf	0463_2024_RO_E.pdf
				Failed URSs:	Failed URSs:	Failed URSs:	Failed URSs:	Failed URSs:	Failed URSs:
				URS07:	URS07:	URS07:	URS07:	URS07:	URS07:

Figure 14 Stability Study: No-Good Documents

Step 4

To increase output yield, it was necessary to work on 2 main directions:

A.   Clean and better structure the CDB in Flex (eq. standardize the name of Manufacturer, Model, etc.….)

B.   Detection improvement of SW tools as:

- Eliminate Case Sensitive
- Ignore Spaces in front of, in the middle or at the end.
- If document contains "-" or "--" and CDB contains "n/a" than results to be pass
- If document contains a table with due date than to verify data, otherwise not
- Add n/a as a possible result as pass or fail

- And many others…..(can be found in CoC Action Tracker file)

In below tables is exemplified are the SW tool output for different SW version and CDB cleaning/adjustments for first two main Calibration Certificate category TESI and MPC

TESI	Verification -v04	Varification -v07	Verification -v08	Verification -v08 updated CDB TESI (URS02-05)	Verification -v09 (change the rule for URS07)	Verification -v09 (update CDB TESI URS06&URS07 and eliminate duplication)	Verification -v09 Estimation after solving URS06 (Thickness Gauges)
Total Tested	1175	1175	1175	1175	1175	1172	1172
Parser Error	114	53	0	0	0	0	0
Fail	1027	961	988	274	578	199	40
Pass	34	161	187	901	597	973	1132
Yield	2.89%	13.70%	15.91%	76.68%	50.81%	83.02%	96.59%

Figure 15 Output category 1 (TESI)

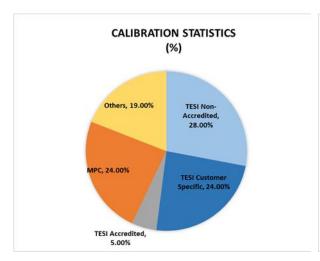| MPC	Verification -v08	Verification -v09 (update CDB MPC) (URS02-05)	Verification -v09 (update CDB URS07)	Verification -v09 (after second correction of URS07 and eliminate duplication)	Verification -v09 Estimated after solving URS06 (AssetID)
Total Tested	1046	1046	1046	1035	1035
Fail	1046	158	96	97	50
Pass	0	888	950	938	985
Yield	0.00%	84.89%	90.82%	90.63%	95.17%

Figure 16 Output for category 2 (MPC)

This step proved to be the most challenging and time-consuming, as Calibration Certificates cover a diverse array of instruments. Not all measurement details provided in the certificates adhere to a standardized template. As a result, certain User Requirement Specifications (URS) for several types of certificates may remain unresolved, potentially leading to "false failures" that could impact the tool's output (yield = ration between Pass/Total Tests)

The fine-tuning of SW tool continues till we reached a Yield >95% which was considered very satisfactory. This means less than 5% of Verified Certificates need to be validated manually by calibration technician.

2.2.3 Cost-Benefit Analysis

Total number of Calibration Certificate that needs to be verified in Flex Timisoara per year is about 3800+ from which 3100+ coming from two Calibration Service Provider, which can be divided in 4 different categories (templates) which are in scope of this project. There are others 700+ Calibration Certificate coming from different sources "Others" which are very unstructured are out of scope for this project in this moment.
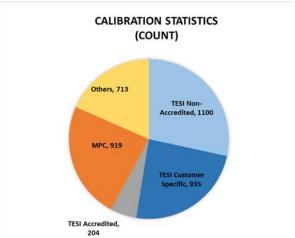
Figure 17 Calibration Certificate statistic (%)   **Figure 18 Calibration Certificate statistic (pcs)**

If we succeed to solve also the 4th category of Calibration Certificate (CC) which means to cover the 3100+ structured Calibration Certificate in scope for this project which means 81% from total Calibration Certificates, considering an average verification time of about 5min/CC the estimated saving per year would be about 258 man-hour (about 32 working days/year).

The cost involved are regarding the further development of HMI and integration of the tool with our local SQL CDB which cannot be estimated at this moment.

Strategic Impact of AI-Driven Calibration Certificate Verification

By developing an intelligent system that leverages the DAI-DSS algorithm to automatically verify the conformity of calibration certificates, organizations unlock a wide range of operational, quality, and sustainability benefits:

- **Ensures Equipment Conformity**: The AI tool validates that manufacturing equipment meets calibration standards, reducing the risk of performance deviations and downtime.
- **Improves Product Quality:** Accurate and timely verification helps prevent quality issues caused by mis-calibrated tools or overlooked certificate errors.
- **Reduces Rework**: Early detection of non-conformities minimizes the need for corrective actions, saving time, resources, and production effort.
- **Lowers Energy Waste:** By avoiding unnecessary rework, the system contributes to more efficient energy usage across production lines.
- **Supports Sustainability Goals**: Less rework and reduced energy consumption translate into a smaller carbon footprint, aligning with environmental targets.
- **Scalable Across Sites and Document Types:** The AI solution can be extended to other plants and adapted to various document formats, promoting standardization and digital transformation across the enterprise.
- **Enhances Workforce Engagement**: Automating repetitive verification tasks allows human employees to focus on more meaningful and intellectually engaging activities—making their roles more attractive while ensuring compliance with industry standards.

2.2.4 Summary of Findings

Even in its "beta" version, the Calibration Certificate Service has already demonstrated its value as a reliable tool for verifying Calibration Certificates. This automated solution aims to significantly decrease review time and minimize the chances of human error, thereby enhancing the accuracy and reliability of the certificate review

process. As a result, it will strengthen internal quality assurance and enable staff who were formerly engaged in manual verification to focus on more value-added tasks.

Main benefits:

- Stable and repeatable.

- High rate of confidence: less than 5% of manual "false call" that need to be verified by human operator.

- Significant time saving.

- Eliminate human errors.

Improvements:

- Developing the HMI to allow for easy selection of the Calibration Certificate type and to display a progress bar (% of completion) when processing multiple files.

- Integration with local SQL CDB

2.3 Assist Decisions about Production Planning (CRF)

2.3.1 Use Case Description

"Use AI to enable human operators to influence a multi-parameter optimization production planning based on personal preferences"

Abstract

In the automotive manufacturing environment, the stamping department plays a critical role in the value chain, as it represents one of the first stages in the transformation of raw materials into components destined for final assembly. In the absence of advanced digital tools for planning and operational management, the organization of production activities and human resources relies on traditional methods, requiring a high degree of coordination, experience, and adaptability from the personnel involved.

Introduction

The production planning process is managed manually, primarily using Excel spreadsheets, printed documentation, and direct communication between production managers, shift supervisors, and operators. Planning begins with the collection of production requirements, which are based on a combination of factors: customer orders (both confirmed and forecasted), stock levels in the warehouse, press capacity, and tool availability.

Based on this input, the production manager defines which part geometries need to be stamped, prioritizing them according to delivery urgency, technical complexity, and the need to optimize tool changeover times. Similar parts are grouped together to minimize the frequency and duration of changes, which are time- and resource-intensive operations.

The quantities to be produced are calculated by considering the hourly capacity of each press, minimum batch sizes for economic efficiency, and the availability of personnel for each shift. The production plan is typically created on a weekly basis, with daily updates to reflect operational changes. Without an MRP (Material Requirements Planning) system, traceability and visibility of priorities depend heavily on interdepartmental communication and manual documentation.

In parallel with production planning, human resources are assigned to the end-of-line stations (Workforce Allocation). Each active press requires a team of one to six operators, depending on the size and complexity of the part. These operators are responsible for unloading the stamped components, performing a basic visual quality check, and placing the parts into the appropriate containers. They do not carry out any additional tasks such as packaging, labelling, or dimensional inspection.

The assignment of operators is managed by the shift supervisor or team leader at the beginning of each shift, based on staff availability, the complexity of the scheduled jobs, and the experience of the workers. In some cases, operators may be rotated between stations to maintain flexibility and reduce fatigue caused by repetitive tasks.

One of the most critical aspects of the current process is the management of unpredictable variables, such as last-minute absences, machine breakdowns, tool non-conformities, or urgent changes in customer orders. In these situations, replanning is carried out in real time through quick decisions made by supervisory personnel, often without the support of digital systems. This results in a strong reliance on the experience and responsiveness of the team, as well as a certain rigidity in resource management.

For example, in the event of an absence, the shift supervisor may decide to temporarily reduce the number of active presses, reassign available operators across multiple stations, or request support from personnel in other departments. These decisions are typically communicated verbally or via whiteboards in the production area and manually updated in production logs.

In summary, the current operating model is based on manual planning and a workforce management approach that depends heavily on the experience and flexibility of the personnel. While this method allows for a certain degree of adaptability to daily operational conditions, it also presents significant limitations in terms of traceability, efficiency, and responsiveness to unforeseen events. These challenges highlight a clear opportunity for the introduction of digital tools to support and optimize decision-making processes.

2.3.2 Use Case Evaluation

In this section the Human-Machine Interface test and evaluation is described. This refers to the evaluation of how the prototypes behave and how the user interface is used and perceived by the testers.

Human Machine Interface (HMI) Validation

2.3.2.1 *Test preparation and execution*

The test preparation phase involved the selection of a group of ten internal volunteer users from CRF, chosen for their familiarity with production processes and typical decision-making activities in industrial contexts. The selected profiles covered a range of competencies, including adaptive systems, production technologies, process integration, and factory sustainability.

In parallel, an evaluation team was established, composed of internal and external experts with backgrounds in ergonomics, industrial processes, and interface design. This team was responsible for defining the test protocol, structuring the tasks to be assigned to users, and preparing the evaluation questionnaire.

The protocol included an introductory phase in which the following were presented:

- the objectives of the FAIRWork project,
- the functionalities of the demonstrator,
- the expected interaction modes,
- the purpose of the test.

Subsequently, each user received a practical demonstration of the tasks to be performed, with the opportunity to take notes. The test was then conducted in two phases: an initial assisted execution and a second autonomous execution, during which both quantitative and qualitative data were collected through the questionnaire.

During the operational phase, each user interacted with the HMI interface following the defined protocol. The assigned tasks covered the entire usage flow of the interface, from initial configuration to production plan generation and result analysis.

The test was structured to evaluate:

- the user's ability to correctly complete each task (quantitative assessment),
- the subjective perception of usability (qualitative assessment using the SUS scale).

Expert personnel observed the interactions, noting any difficulties, doubts, or recurring behaviors. At the end of each session, the user completed the second part of the questionnaire, providing direct feedback on the user experience.

All collected data were subsequently analyzed to identify strengths, critical issues, and opportunities for improvement of the interface.

2.3.2.2   Questionnaire description

The image below shows the questionnaire prepared for the test and submitted to the participants:

Figure 19 HMI evaluation questionnaire

The FAIRWork HMI validation questionnaire was structured into two sections: an operational part and a user perception part.

The first section, called "quantitative test", focused on operational tasks to measure the effectiveness of the interface.

The second section collected subjective feedback from volunteers regarding their experience using the interface. Although this phase is qualitative in terms of data collection (based on personal perceptions), the evaluation was carried out using the System Usability Scale (SUS) method, which provides a quantitative score of perceived usability.

The expression "quantitative test (SUS)" is used frequently throughout this document and may appear ambiguous.

The System Usability Scale (SUS) collects subjective feedback from users, but translates it into a standardized numerical score, enabling a quantitative evaluation of perceived usability.

Therefore, while the data collection is qualitative in nature (based on personal opinions), the SUS method allows for a quantification of qualitative data, which justifies the use of the term in the context of this document.

The overall objective is to assess the effectiveness, usability, and intuitiveness of the interface developed for the Production Planning and Resource Allocation Service.

Part 1: Quantitative Test

The quantitative test is based on a sequence of tasks that reflect the main functionalities of the HMI. Each task was designed to simulate a real action that a user would perform when using the system in a production environment.

- Task 1. Application Launch: The user accesses the interface via a web link, which opens the system's start screen. From here, data can be uploaded, and planning parameters configured.
- Task 2. Order File Upload: The user selects an Excel file containing the orders to be scheduled. Each row represents a geometry to be produced, with information on quantity, priority, delivery date, and required number of molds.
- Task 3. Planning Parameter Setup: The user defines the time window (start date and number of days to plan) and the type of personnel availability (scheduled or actual), allowing simulation of both theoretical and real scenarios.
- Task 4. Personnel Availability Selection: The user chooses whether to use scheduled attendance data (standard shifts) or actual data (from badge system). This choice directly affects the feasibility of the plan.
- Task 5. Order Allocation Method Configuration: The interface allows selection among different optimization approaches (e.g., Constraint Programming), with varying objectives: balancing delays and machine usage, minimizing tardiness, or minimizing makespan.
- Task 6. Worker Allocation Method Configuration: The user can choose among several algorithms (Constraint Programming, Reinforcement Learning, Monte Carlo Tree Search) and assignment criteria (experience, preferences, resilience, balance).
- Task 7. AI Service Activation: By pressing the "Trigger AI Service" button, the system generates a production and worker allocation plan. Processing takes a few seconds and returns an interactive visualization of the results.
- Task 8. Planning Results Analysis: The interface displays results graphically, with performance indicators (experience, preferences, resilience, transparency) and a timeline of activities for each production line.
- Task 9. Saving the Generated Plan: The user can save the plan in Excel format, containing all information related to orders, lines, assigned workers, and any operational alerts.
- Task 10. Interpretation of Operational Alerts: If the system detects critical issues (e.g., insufficient number of workers), specific alerts are shown. The user must be able to interpret them and decide whether to intervene manually.

Part 2: Qualitative Test (SUS)

The second part of the questionnaire is based on the SUS standard, consisting of ten statements rated on a Likert scale. The questions explore the user's subjective perception regarding ease of use, consistency, feature integration, need for support, and confidence in the system. The statements are:

- Q1. "I think that I would like to use this software frequently"
- Q2. "I found this software unnecessarily complex."
- Q3. "I thought this software was easy to use."
- Q4. "I think that I would need assistance to be able to use this software."

- Q5. "I found the various functions in this software were well integrated."
- Q6. "I thought there was too much inconsistency in this software."
- Q7. "I would imagine that most people would learn to use this software very quickly."
- Q8. "I found this software very cumbersome/awkward to use."
- Q9. "I felt very confident using this software."
- Q10. "I needed to learn a lot of things before I could get going with this software."

The final score provides a concise indication of perceived usability, useful for comparisons with other systems or versions.

2.3.2.3 Quantitative test results

The quantitative test assessed the users' ability to correctly complete 10 operational tasks. Each task was scored as follows:

- 1 for correct completion,
- 0.5 for partial completion,
- 0 for failure to complete.

The results show a high level of interface usability among users. Individual averages ranged from 80% to 100%, with an overall average of 89%, indicating a very good usability level according to standard evaluation metrics

Quantitative test	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	Total	St. Dev.
Task 1	1	1	1	1	1	1	1	1	1	1	**100%**	0
Task 2	1	1	1		1	1	1	1	1	1	**100%**	0
Task 3	1	1	1		1	0,5	1	0,5	1	1	**90%**	0,21
Task 4	1	0,5	0,5		1	1	1	1	1	0,5	**85%**	0,24
Task 5	0,5	1	1	0,5	1	0,5	1	1	0,5	1	**80%**	0,26
Task 6	0,5	1	1	0,5	1	1	1	1	0,5	1	**85%**	0,24
Task 7	1	1	1		1	1	1	1	1	1	**100%**	0
Task 8	1	0,5	0,5	0,5	1	0,5	0,5	1	1	0,5	**70%**	0,26
Task 9	1	1	1		1	1	1	1	1	1	**100%**	0
Task 10	0,5	1	1	0,5	1	1	1	0,5	0,5	1	**80%**	0,26
Result: 89%	**85%**	**90%**	**90%**	**80%**	**100%**	**85%**	**95%**	**90%**	**85%**	**90%**		

Figure 20 HMI evaluation, quantitative test results

Despite the strong overall performance, some tasks received lower scores. A detailed analysis highlighted a few critical areas:

- Tasks 3 and 4 (planning parameter configuration) showed uncertainty related to understanding time settings and staff availability options.
- Tasks 5 and 6 (allocation method configuration) revealed difficulties in selecting between algorithms and criteria, suggesting a need for clearer explanations or informational support.
- Tasks 8 and 10 (results analysis and alert interpretation) were the most complex, indicating that reading and interpreting AI-generated KPIs and timelines requires a more guided or explanatory interface.

Standard deviation analysis confirmed these findings:
- The most intuitive tasks (1, 2, 7, 9) were completed correctly by all users.
- Tasks with the highest variability (5, 8, 10) represent the areas where user experience improvements are most urgently needed.
- Tasks with intermediate variability (3, 4, 6) were generally understandable but could benefit from additional support.

In summary, the quantitative test confirms the robustness of the interface but suggests targeted improvements to enhance the clarity of more complex features and better support users in advanced decision-making steps.

Below are some comments collected during the FAIRWork HMI test, based on the quantitative results, and identified critical areas:

"I wasn't sure what time range to enter for planning. Maybe an example or automatic suggestion would help.": Related to Tasks 3 and 4, highlighting uncertainty in configuring time parameters.

"Choosing between the different allocation algorithms isn't straightforward. A short description next to each option would be helpful.": Refers to Tasks 5 and 6, where the logic behind selecting CP, RL, or MCTS was not always clear.

"The result visualization is interesting, but I struggled to understand what the KPIs meant at first.": A typical comment for Task 8, which requires interpretation of complex indicators.

"I expected a clearer message when there were issues with operator assignments. The alert is not very visible.": Linked to Task 10, regarding the interpretation of operational warnings.

"The initial operations are very simple, but when it comes to analysis, you really need to know the system well.": A general reflection summarizing the difference between intuitive and more advanced tasks.

2.3.2.4 Qualitative test (SUS) results

The System Usability Scale (SUS) provided a subjective assessment of perceived usability. The score was calculated using the standard formula:

- For odd-numbered questions: score = rating - 1
- For even-numbered questions: score = 5 - rating
- Sum of the results × 2.5

The SUS scores obtained from the 10 users ranged from 65% to 95%, with an overall average of 84%. This value corresponds to an "excellent" rating according to the SUS reference scale, indicating high user satisfaction and a strong willingness to adopt the tool.

Qualitative test (SUS)	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	Total	St. Dev.
Query 1	5	4	4	4	5	4	4	5	4	4	**83%**	0,48
Query 2	2	2	1	2	1	2	1	2	2	1	**85%**	0,52
Query 3	4	5	5	4	5	4	5	5	4	5	**90%**	0,52
Query 4	2	1	1	3	1	2	1	1	3	1	**85%**	0,84
Query 5	4	3	4	3	4	4	4	5	4	5	**75%**	0,67
Query 6	2	1	1	2	1	2	1	1	2	2	**88%**	0,53
Query 7	3	5	5	3	5	4	5	5	4	5	**85%**	0,84
Query 8	2	1	1	2	1	2	1	1	2	1	**90%**	0,52
Query 9	4	5	4	3	4	4	5	5	4	4	**80%**	0,63
Query 10	2	2	1	2	2	1	2	2	1	2	**83%**	0,48
Result: 84%	**75%**	**88%**	**93%**	**65%**	**93%**	**78%**	**93%**	**95%**	**75%**	**90%**		

Figure 21 HMI evaluation, qualitative test (SUS) results

Despite the average score of 84% indicating excellent perceived usability, some responses suggest areas for improvement. The following statements received relatively lower scores or highlighted potential issues:

- Function integration (Q5 - 75%): Some users perceived a lack of coherence between interface modules. This suggests the need to improve the fluidity of operational flows and visual consistency.
- Confidence in use (Q9 - 80%): Although the score is good, some users did not feel fully confident interacting with the system, likely due to the complexity of certain advanced features.
- Learning curve (Q10 - 83%): Initial learning was perceived as challenging by some participants, suggesting the opportunity to introduce more structured onboarding tools (e.g., tutorials, guided tours).

For completeness, the following statements received the highest scores:

- Q3 - "I thought this software was easy to use": 90%
- Q8 - "I found this software very cumbersome to use" (inverted): 90%
- Q6 - "I thought there was too much inconsistency in this software" (inverted): 88%

These values confirm that, despite some areas for improvement, the overall perceived usability is very good.

Moreover, high standard deviations for some questions (Q4 and Q7) indicate significant variability in user experiences, likely related to users' technical backgrounds. Some felt autonomous, while others perceived a need for assistance.

Conversely, questions with low variability (Q1, Q2, Q3, Q8, Q10) show a shared perception on fundamental aspects such as ease of use and initial understanding.

In summary, the SUS test confirms that the interface is well-designed and appreciated by users, but also highlights the importance of:

• improving integration between functionalities,
• strengthening user support during initial use,
• increasing transparency and feedback during interaction.

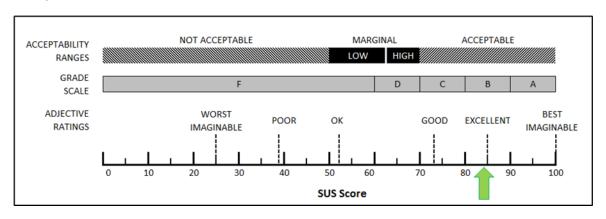The image below shows the result obtained on the SUS test evaluation scale

Figure 22 SUS test evaluation scale

Below are some comments collected during the FAIRWork HMI test, based on the qualitative test results and identified critical areas:

"Overall, it's easy to use, but at first I needed some time to get oriented with the various functions.": Reflection on the initial learning curve (Q10).

"Some screens feel disconnected, as if there's no logical flow between functions.": Comment related to the perception of poor integration between functionalities (Q5).

"I'd like to have a short tutorial the first time I log in, to immediately understand what to do.": Suggestion to improve onboarding and reduce the need for assistance (Q4, Q10).

"Once I understood how it works, it's convenient. But at first, I wasn't confident I was making the right choices.": Related to confidence in using the system (Q9).

"It positively surprised me: I thought it would be more complicated, but it's actually quite intuitive.": Positive comment reflecting high scores on ease of use (Q3) and comfort (Q8).

CRF - DECISION-MAKING MODEL VALIDATION

2.3.2.5   Test preparation and execution

The validation of the decision-making model was conducted during week 26 of 2025, in accordance with the methodology outlined in Section 5.3 of Deliverable D5.3. The objective was to evaluate the system's ability to propose effective and context-aware decisions for production planning and workforce allocation, under realistic operational conditions.

The preparation phase began with the identification of key personnel from the Mirafiori Press Shop (assisted by CRF), selected for their in-depth knowledge of planning processes, resource management, and operational constraints. The validation team was coordinated by the Production Planning & Control Manager, who designated specialists from the areas of Handling Operations, Lines and Materials Planning, and Manpower Analysis to participate in the test sessions.

The validation process was structured into three main phases:

1. Informative Session: Participants were introduced to the FAIRWork project, the concept of democratic decision-making, and the functionalities of the DAI-DSS system. Emphasis was placed on the fairness, adaptability, and transparency of the proposed solutions.
2. Formative Session: A hands-on training session was conducted to familiarize users with the interface and the underlying logic of the decision-making model. Participants were guided through the input requirements (e.g., order files, availability data), configuration parameters, and interpretation of results. The session also encouraged experimentation with different planning scenarios to observe the system's behavior under varying conditions.
3. Operational Test: Participants then carried out several test sessions, planning multiple weeks of production while repeatedly varying the parameters.

The questionnaire used is essentially the same as the one presented in Deliverable D5.3; it is reproduced here for convenience and ease of reading.
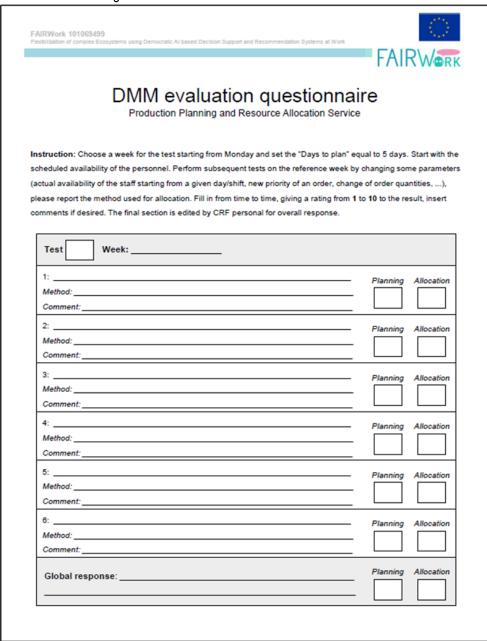

Figure 23 Decision-Making Model evaluation questionnaire

To ensure consistency and minimize bias, each session was facilitated by CRF personnel with expertise in the use of the demonstrator. This allowed participants to focus on the decision logic and outcomes, rather than on interface navigation.

Regarding the Operational Test, in the initial planning phase, as outlined in document D.5.3 (Section 5.3.2), a structured protocol was defined for the execution of the test sessions. The expected procedure was as follows:

- The candidate would select a reference week for the test, starting from Monday.
- The "Days to plan" parameter would be set to 5 days.
- The test would begin using the scheduled availability of the personnel.
- The candidate would document the method used for resource allocation.
- During the planning process, the candidate would assign a score from 1 to 10 to the outcome of each planning attempt.
- Optional comments could be added to justify the score or provide qualitative feedback.
- The candidate would then perform additional tests on the same reference week by modifying certain parameters (e.g., actual staff availability from a specific day/shift, priority changes for an order, variation in order quantities, etc.).
- At the end of each session, CRF personnel would complete the final section of the questionnaire, calculating the average score and adding a final comment.
- This procedure was to be repeated for at least 10 different weeks, resulting in a corresponding number of completed questionnaires. It was estimated that one day would be sufficient to complete all test sessions.

Although the initial test sessions adhered closely to the protocol defined in Section 5.3.2 of D.5.3, as the validation progressed, it became increasingly challenging to follow the procedure rigorously. In particular, the rapid succession of simulations and frequent parameter changes made it impractical to complete the evaluation questionnaires in full detail for each planning attempt. One of the most critical aspects that proved difficult to maintain was the assignment of a score from 1 to 10 after each planning iteration, as it was often challenging to track and annotate all the changes made during the session in a structured manner.

This deviation from the protocol is interpreted positively: the candidates demonstrated a high level of engagement and curiosity, actively exploring how different parameter configurations influenced the planning outcomes. Rather than limiting the analysis to 10 predefined weeks, the validation was extended to include all available weeks (16), thereby enriching the dataset and the overall evaluation.

At the conclusion of each test session, only the final section of the questionnaire was completed. This included an overall assessment for the week and user comments summarizing their experience and observations. During the process, several fundamental issues related to the demonstrator under evaluation emerged. These will be discussed in detail in the following sections of this document. Once identified, these issues were acknowledged and considered only once in the evaluation, as they are expected to be addressed in future versions of the demonstrator.

The test protocol also included the measurement of key performance indicators (KPIs), as defined in D5.3 and D4.1.1, such as:

- Decision Time: time required to generate a valid plan once all input data were available.
- Alternative Plans: number of viable solutions proposed by the system.
- Planning and Allocation Scores: average ratings assigned by users.
- Operational Feedback: comments on usability, transparency, and adaptability of the model.

Preliminary results indicate that, despite the identified limitations, the overall user feedback was positive, confirming the system's potential for real-world deployment.

2.3.2.6 Production Planning analysis

In this section, the analysis criteria adopted for evaluating the Decision-Making Model will be presented, with specific reference to the "Production Planning" aspect. The developed interface is at times somewhat difficult to interpret, as it contains a large amount of information and does not provide an immediate and easily understandable visual response.

As an example, a detailed analysis of week 36 is provided below, starting from September 4th with a time horizon of 7 days and the "Availability Type" set to "Programmed". The two allocation methods for orders and workers are set to "Constraint Programming - balanced".

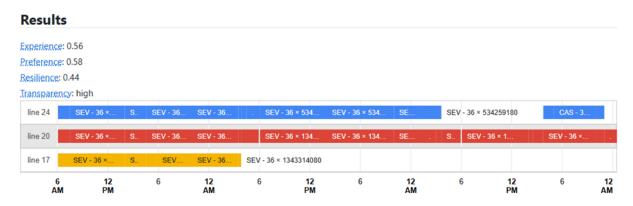The Gantt chart generated by the demonstrator is shown below.

Figure 24 Production Gantt week 36, CP-balanced

By downloading the corresponding file using the "Save Assignment" function and sorting the rows by "Line" and "Start", the result for line 17 is as follows:

2	Task	Start	Finish	Production line	Setup time indicator	Worker assignments
3	SEV - 36 x 531359140	04-09 06:00	04-09 07:15	line 17	VERO	15004479 15016633 15020309 15028837
4	SEV - 36 x 531359140	04-09 07:15	04-09 14:00	line 17	FALSO	15004479 15016633 15020309 15028837
5	SEV - 36 x 531359140	04-09 14:00	04-09 16:35	line 17	FALSO	15015125 15015264 15015351 15016591
6	SEV - 36 x 1343314080	04-09 16:35	04-09 17:30	line 17	VERO	15013367 15014964 15015264 15015351 15028914
7	SEV - 36 x 1343314080	04-09 17:30	04-09 17:50	line 17	VERO	15013367 15014729 15015264 15015568 15028914 15040627
8	SEV - 36 x 1343314080	04-09 17:50	04-09 22:00	line 17	FALSO	15013367 15014729 15015264 15015568 15028914 15040627
9	SEV - 36 x 1343314080	04-09 22:00	05-09 03:50	line 17	FALSO	15015554

Figure 25 Production Gantt week 36 (line 17), CP-balanced

In detail:

- Row 3: This is a mold change lasting 1 hour and 15 minutes, to start producing geometry 531359140 from the beginning of the first shift at 6:00 AM (in bold). Worker assignment is planned, but, it is not necessary because the line does not produce during setup.
- Row 4: Production proceeds correctly until the end of the shift.
- Row 5: Production resumes in the second shift (starting at 2:00 PM, in bold) with the change of workers scheduled for the second shift.
- Rows 6–7: From 4:35 PM to 5:50 PM (i.e., for 1 hour and 15 minutes), there is a mold change to switch to geometry 1343314080. This operation is split into two rows (which generate two blocks on the Gantt chart). This is unnecessary and is likely due to the change in worker assignment, but as mentioned, no workers are required during line setup.
- Row 8: Production resumes with the new geometry until the end of the shift (at 10:00 PM, in bold).
- Row 9: Starting at 10:00 PM, with the third shift (in bold), production resumes.

The result for line 20 is as follows:

	Task	Start	Finish	Production line	Setup time indicator	Worker assignments
3	SEV - 36 x 1342266080	04-09 06:00	04-09 07:30	line 20	VERO	15013130 15013533 15014209 15014212
4	SEV - 36 x 1342266080	04-09 07:30	04-09 14:00	line 20	FALSO	15013130 15013533 15014209 15014212
5	SEV - 36 x 1342266080	04-09 14:00	04-09 16:35	line 20	FALSO	15014729 15015568 15017049 15040627
6	SEV - 36 x 1342266080	04-09 16:35	04-09 17:30	line 20	FALSO	15014729 15015568 15017049 15040627
7	SEV - 36 x 1342236080	04-09 17:30	04-09 18:30	line 20	VERO	15009882 15015125 15015351 15028961
8	SEV - 36 x 1342236080	04-09 18:30	04-09 22:00	line 20	FALSO	15009882 15015125 15015351 15028961
9	SEV - 36 x 1342236080	04-09 22:00	05-09 03:50	line 20	FALSO	15015650
10	SEV - 36 x 1342236080	05-09 03:50	05-09 05:10	line 20	FALSO	15015650
11	SEV - 36 x 1342236080	05-09 05:10	05-09 06:00	line 20	FALSO	
12	SEV - 36 x 1342236080	05-09 06:00	05-09 12:25	line 20	FALSO	15004479 15015652 15015653 15016633
13	SEV - 36 x 1342236080	05-09 12:25	05-09 14:00	line 20	FALSO	15004479 15015652 15015653 15016633
14	SEV - 36 x 1342236080	05-09 14:00	05-09 15:30	line 20	FALSO	15009882 15015125 15015351 15028961
15	SEV - 36 x 534259080	05-09 15:30	05-09 17:00	line 20	VERO	15004696 15009882 15013367 15014964
16	SEV - 36 x 534259080	05-09 17:00	05-09 22:00	line 20	FALSO	15004696 15009882 15013367 15014964
17	SEV - 36 x 534259080	05-09 22:00	06-09 06:00	line 20	FALSO	15004666 15015349 15015510
18	SEV - 36 x 534259080	06-09 06:00	06-09 13:00	line 20	FALSO	15004479 15009935 15013533 15013870
19	SEV - 36 x 1343327080	06-09 13:00	06-09 13:58	line 20	VERO	15013130 15014209 15015653 15020309
20	SEV - 36 x 1343327080	06-09 13:58	06-09 14:00	line 20	VERO	15013130 15015652 15015653 15020309
21	SEV - 36 x 1343327080	06-09 14:00	06-09 14:30	line 20	VERO	15009882 15015568 15017049 15028914
22	SEV - 36 x 1343327080	06-09 14:30	06-09 22:00	line 20	FALSO	15009882 15015568 15017049 15028914
23	SEV - 36 x 1343327080	06-09 22:00	07-09 00:30	line 20	FALSO	15015349 15015554 15015650

Figure 26 Production Gantt week 36 (line 20), CP-balanced

By applying the same analysis method extensively explained for line 17, several aspects are highlighted for line 20:

- Rows 5–6: The task is split into two blocks, but the reason is unclear, especially since the assigned workers do not change.
- Rows 9–10–11: The task is divided into three blocks, with the first two being identical. Resource 15015650 is moved to line 24 at 5:10 AM on September 5th for a mold change (an unnecessary reassignment). Moreover, during the entire third shift, there is a significant shortage of workers on line 20. It is recommended to set a limit on the number of missing workers, beyond which the line should not be activated.
- Rows 12–13: The task is split into two blocks, but there is no difference between them.
- Rows 19–20–21: This is a line setup change; therefore, it could be a single task or, at most, split into two due to the shift change. Leaving aside this already discussed aspect, upon investigating the system logic, the first task split is due to a change in one of the four workers. It is preferable not to change a worker's role within the same shift, especially since the second block lasts only two minutes. Furthermore, this worker is not reassigned, and the other two lines are idle after 1:58 PM on September 6th.

The following section pertains to the final line under analysis, the number 24:

Production Planning analysis

	Task	Start	Finish	Production line	Setup time indicator	Worker assignments
3	SEV - 36 x 534259180	04-09 06:00	04-09 07:30	line 24	VERO	15013870 15017042 15028786 15097932
4	SEV - 36 x 534259180	04-09 07:30	04-09 14:00	line 24	FALSO	15013870 15017042 15028786 15097932
5	SEV - 36 x 534259180	04-09 14:00	04-09 16:35	line 24	FALSO	15009882 15013367 15015261 15028914
6	SEV - 36 x 534259180	04-09 16:35	04-09 17:30	line 24	FALSO	15009882 15015125 15015261 15016591
7	SEV - 36 x 534259180	04-09 17:30	04-09 22:00	line 24	FALSO	15014964 15015261 15016591 15017049
8	SEV - 36 x 534259180	04-09 22:00	05-09 03:50	line 24	FALSO	15015349
9	SEV - 36 x 534259180	05-09 03:50	05-09 05:10	line 24	FALSO	15015349 15015554
10	CAS - 36 x 505597580	05-09 05:10	05-09 06:00	line 24	VERO	15015349 15015554 15015650
11	CAS - 36 x 505597580	05-09 06:00	05-09 06:25	line 24	VERO	15013130 15013533 15014209 15017042
12	CAS - 36 x 505597580	05-09 06:25	05-09 12:25	line 24	FALSO	15013130 15013533 15014209 15017042
13	SEV - 36 x 533908540	05-09 15:30	05-09 17:00	line 24	VERO	15014729 15015125 15015261 15017049
14	SEV - 36 x 533908540	05-09 17:00	05-09 22:00	line 24	FALSO	15014729 15015125 15015261 15017049
15	SEV - 36 x 533908540	05-09 22:00	06-09 06:00	line 24	FALSO	15015554 15015650
16	SEV - 36 x 533908540	06-09 06:00	06-09 13:00	line 24	FALSO	15013130 15015652 15020309 15028786
17	SEV - 36 x 533908540	06-09 13:00	06-09 13:58	line 24	FALSO	15015652 15016633 15017042 15028786

Figure 27 Production Gantt week 36 (line 24), CP-balanced

The main observations are as follows:

- Rows 5–6–7: The task is split into three blocks due to a change during the shift in the personnel assigned to unloading operations. As previously discussed, it is preferable not to change personnel assignments within the same shift.
- Rows 8–9: Same as the point above.
- Rows 10–11: This is a line setup change; it could be considered a single task, but the current split is acceptable given the shift change.
- Row 12-13: A 3:05 gap is observed in the scheduling between the two tasks. The reason for this pause is unclear.
- Rows 16–17: Same considerations as for rows 5–6–7.

The analysis was repeated by changing the Order Allocation Method to "Constraint Programming - tardiness"
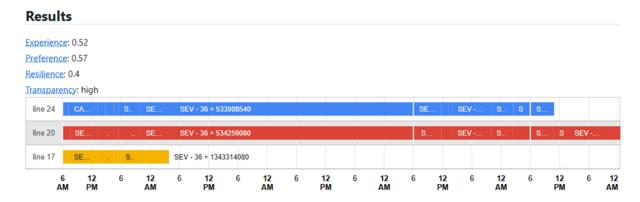

Figure 28 Production Gantt week 36, CP-tardiness

In this case, as indicated in the demonstrator description the scheduling goal is to minimize tardiness defined as the time by which an order misses its deadline. Make span is not considered in this optimization. As a result, the system maximizes on-time deliveries, even if it reduces machine utilization by prioritizing deadline adherence over throughput.

As an example, consider line 20, shown below:

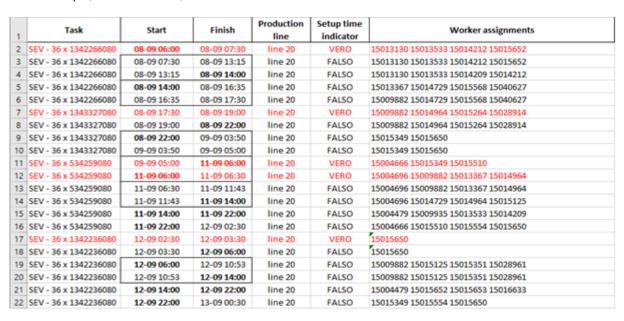| 1	Task	Start	Finish	Production line	Setup time indicator	Worker assignments
2	SEV - 36 x 1342266080	08-09 06:00	08-09 07:30	line 20	VERO	15013130 15013533 15014212 15015652
3	SEV - 36 x 1342266080	08-09 07:30	08-09 13:15	line 20	FALSO	15013130 15013533 15014212 15015652
4	SEV - 36 x 1342266080	08-09 13:15	08-09 14:00	line 20	FALSO	15013130 15013533 15014209 15014212
5	SEV - 36 x 1342266080	08-09 14:00	08-09 16:35	line 20	FALSO	15013367 15014729 15015568 15040627
6	SEV - 36 x 1342266080	08-09 16:35	08-09 17:30	line 20	FALSO	15009882 15014729 15015568 15040627
7	SEV - 36 x 1343327080	08-09 17:30	08-09 19:00	line 20	VERO	15009882 15014964 15015264 15028914
8	SEV - 36 x 1343327080	08-09 19:00	08-09 22:00	line 20	FALSO	15009882 15014964 15015264 15028914
9	SEV - 36 x 1343327080	08-09 22:00	09-09 03:50	line 20	FALSO	15015349 15015650
10	SEV - 36 x 1343327080	09-09 03:50	09-09 05:00	line 20	FALSO	15015349 15015650
11	SEV - 36 x 534259080	09-09 05:00	11-09 06:00	line 20	VERO	15004666 15015349 15015510
12	SEV - 36 x 534259080	11-09 06:00	11-09 06:30	line 20	VERO	15004696 15009882 15013367 15014964
13	SEV - 36 x 534259080	11-09 06:30	11-09 11:43	line 20	FALSO	15004696 15009882 15013367 15014964
14	SEV - 36 x 534259080	11-09 11:43	11-09 14:00	line 20	FALSO	15004696 15014729 15014964 15015125
15	SEV - 36 x 534259080	11-09 14:00	11-09 22:00	line 20	FALSO	15004479 15009935 15013533 15014209
16	SEV - 36 x 534259080	11-09 22:00	12-09 02:30	line 20	FALSO	15004666 15015510 15015554 15015650
17	SEV - 36 x 1342236080	12-09 02:30	12-09 03:30	line 20	VERO	15015650
18	SEV - 36 x 1342236080	12-09 03:30	12-09 06:00	line 20	FALSO	15015650
19	SEV - 36 x 1342236080	12-09 06:00	12-09 10:53	line 20	FALSO	15009882 15015125 15015351 15028961
20	SEV - 36 x 1342236080	12-09 10:53	12-09 14:00	line 20	FALSO	15009882 15015125 15015351 15028961
21	SEV - 36 x 1342236080	12-09 14:00	12-09 22:00	line 20	FALSO	15004479 15015652 15015653 15016633
22	SEV - 36 x 1342236080	12-09 22:00	13-09 00:30	line 20	FALSO	15015349 15015554 15015650

Figure 29 Production Gantt week 36 (line 20), CP-tardiness

The concept of 'tardiness' was misinterpreted: production starts on the due date itself (September 8th), whereas it should be completed by the due date. The same issue applies to the other two lines as well. Nevertheless, the system's flexibility is commendable, as it can recalculate a schedule by shifting production forward. Acknowledging that this is the first version of the demonstrator, we suggest as a future improvement the modification of the algorithm so that production occurs 'close to the due date.'

Regarding the figure shown above, the same observations previously discussed are roughly highlighted once again.

The analysis was repeated by changing the Order Allocation Method to "Constraint Programming - make span ":
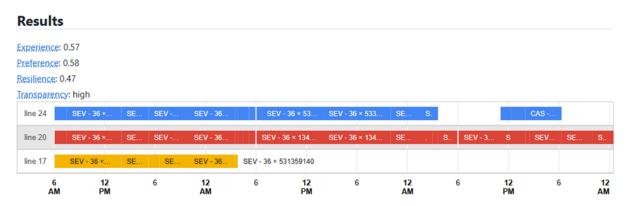
Figure 30 Production Gantt week 36, CP-make span

As reported in the demonstrator description, orders are assigned to production lines in a way that minimizes the make span, that is the total time required to complete all orders. This service does not consider delivery deadlines. Its sole focus is on minimizing make span, resulting in high machine utilization, even if it means missing some deadlines in favor of throughput efficiency.

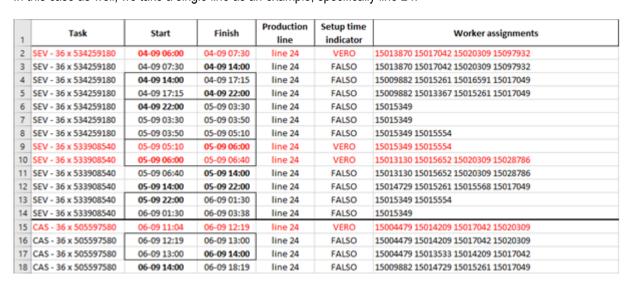In this case as well, we take a single line as an example, specifically line 24:

	Task	Start	Finish	Production line	Setup time indicator	Worker assignments
2	SEV - 36 x 534259180	04-09 06:00	04-09 07:30	line 24	VERO	15013870 15017042 15020309 15097932
3	SEV - 36 x 534259180	04-09 07:30	04-09 14:00	line 24	FALSO	15013870 15017042 15020309 15097932
4	SEV - 36 x 534259180	04-09 14:00	04-09 17:15	line 24	FALSO	15009882 15015261 15016591 15017049
5	SEV - 36 x 534259180	04-09 17:15	04-09 22:00	line 24	FALSO	15009882 15013367 15015261 15017049
6	SEV - 36 x 534259180	04-09 22:00	05-09 03:30	line 24	FALSO	15015349
7	SEV - 36 x 534259180	05-09 03:30	05-09 03:50	line 24	FALSO	15015349
8	SEV - 36 x 534259180	05-09 03:50	05-09 05:10	line 24	FALSO	15015349 15015554
9	SEV - 36 x 533908540	05-09 05:10	05-09 06:00	line 24	VERO	15015349 15015554
10	SEV - 36 x 533908540	05-09 06:00	05-09 06:40	line 24	VERO	15013130 15015652 15020309 15028786
11	SEV - 36 x 533908540	05-09 06:40	05-09 14:00	line 24	FALSO	15013130 15015652 15020309 15028786
12	SEV - 36 x 533908540	05-09 14:00	05-09 22:00	line 24	FALSO	15014729 15015261 15015568 15017049
13	SEV - 36 x 533908540	05-09 22:00	06-09 01:30	line 24	FALSO	15015349 15015554
14	SEV - 36 x 533908540	06-09 01:30	06-09 03:38	line 24	FALSO	15015349
15	CAS - 36 x 505597580	06-09 11:04	06-09 12:19	line 24	VERO	15004479 15014209 15017042 15020309
16	CAS - 36 x 505597580	06-09 12:19	06-09 13:00	line 24	FALSO	15004479 15014209 15017042 15020309
17	CAS - 36 x 505597580	06-09 13:00	06-09 14:00	line 24	FALSO	15004479 15013533 15014209 15017042
18	CAS - 36 x 505597580	06-09 14:00	06-09 18:19	line 24	FALSO	15009882 15014729 15015261 15017049

Figure 31 Production Gantt week 36 (line 24), CP-makespan

Compared to the "balanced" order allocation method, in this case production ends slightly later, at 18:19 instead of 13:58 on the same day. This delay is due to a break, which is also difficult to interpret, lasting 7 hours and 26 minutes, from 03:38 to 11:04. Aside from this, the same observations previously noted are roughly applicable here as well.

For completeness, the Gantt charts of the other order allocation methods are also provided. Below is the one for the "Optimize – balanced":
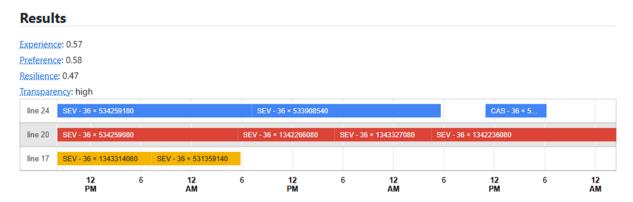

Figure 32 Production Gantt week 36, Optimize-balanced

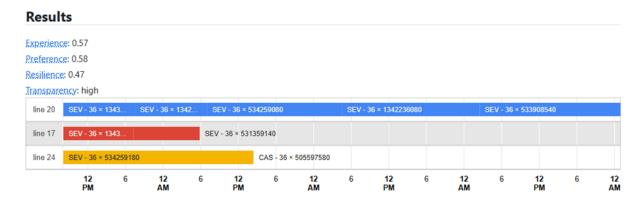Therefore, for the "Optimize - tardiness":

Figure 33 Production Gantt week 36, Optimize-tardiness

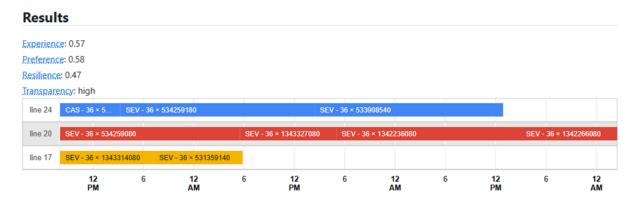Finally, for the last one, the "Optimize - makespan":

Figure 34 Production Gantt week 36, Optimize-makespan

It should be noted that the last three methods, currently under validation, were used less frequently. It is commendable that the system allows production planning alone, without considering resource allocation; however, this aspect was intentionally considered.

As for the first three methods, the preferred one turned out to be the 'balanced' method. The other two were also tested during the experiment, but in a more limited way. This is mainly due to a fundamental misinterpretation in

the 'tardiness' method, while the 'makespan' method yields results that are roughly equivalent to those of the 'balanced' method.

Further structured analyses are described below, covering all 16 weeks, with a time horizon of 7 days each, and the "Availability Type" set to "Programmed". The two allocation methods for orders and workers are set to "Constraint Programming - balanced".

Orders not fully fulfilled: only in two cases were the orders not completely fulfilled, although this is due to a bug in the model:

- SEV-40 x 533908540: scheduled in week 40 with a target of 6,290 units, the plan transitions from a task with a cumulative quantity of 4,350 units to one of just 10 units, following an additional production of 1,950 units. This results in a total of 6,300 units—10 units over the target.
- SEV-44 x 1342236080: scheduled in week 44 with a target of 6,300 units, the plan moves from a task with a cumulative quantity of 4,725 units to one of 3,000 units, after an additional production of 4,575 units. This leads to a total of 9,300 units—3,000 units above the target.

In both cases, the issue appears to be a bug in the model, where only the excess portion beyond the target is considered once the target is exceeded.

Unjustified Pauses on Production Lines: following the previously identified pause during week 36, an investigation was conducted to determine whether this was an isolated incident or a recurring behavior of the scheduling model.

The analysis across all weeks revealed only two occurrences:

- Week 36: Order SEV-36 x 533908540 scheduled on line 24 shows a gap between 12:25 and 15:30 (duration: 03:05).
- Week 40: Order SEV-40 x 1343327080 scheduled on line 20 shows a gap between 15:30 and 15:38 (duration: 00:08).

The conclusion is that the behavior appears to be random and minimal, likely caused by another minor bug in the system.

Production Scheduled Too Early Compared to Due Dates: it has been observed that the model tends to schedule production completion significantly earlier than the actual due dates. While this behavior is certainly cautious, it is not always necessary and may lead to overloading of finished goods storage areas.

Typically, production starts on the same day the simulation is launched. As previously discussed, a scheduling approach that aligns more closely with the due dates is often preferred.

Although this can currently be managed by adjusting the simulation start date in the initial configuration, a future enhancement could automate this behavior, optimizing warehouse usage without compromising delivery reliability.

An analysis was conducted on a total of 128 production tasks to evaluate how far in advance they are completed relative to their respective due dates. The results show a clear tendency of the model to schedule production well ahead of deadlines. While this approach may be conservative and ensure timely delivery, it can also lead to unnecessary early stockpiling of finished goods.

The distribution of lead times is as follows:

- Less than 1 day: 5 tasks
- Between 1 and 2 days: 7 tasks
- Between 2 and 3 days: 26 tasks

- Between 3 and 4 days: 58 tasks
- Between 4 and 5 days: 32 tasks

These findings suggest that many tasks (over 90%) are completed more than two days in advance, with a significant concentration between 3 and 5 days. This behavior, while safe, may be optimized in future model versions to better balance production timing and warehouse efficiency.

Resource Allocation with warnings given that component production is typically completed ahead of the scheduled due date, an analysis was conducted to assess how often the model proposes a problematic resource allocation. Specifically, the focus was on identifying how many times a valid task (i.e., not a line setup) includes the warning "Too few workers assigned".

The following table summarizes the results:

Week (from Monday)	Warnings	Valid tasks	Percentage
Week 36	9	30	30.0%
Week 37	5	27	18.5%
Week 38	11	33	33.3%
Week 39	8	24	33.3%
Week 40	7	25	28.0%
Week 41	12	31	38.7%
Week 42	11	28	39.3%
Week 43	11	35	31.4%
Week 44	13	24	54.2%
Week 45	7	24	29.2%
Week 46	3	25	12.0%
Week 47	7	18	38.9%
Week 48	6	18	33.3%
Week 49	7	19	36.8%
Week 50	6	25	24.0%
Week 51	6	21	28.6%

Figure 35 Resource Allocation with warnings

It is important to note that this is not considered a model error. On the contrary, the possibility of generating such warnings was explicitly stated as part of the model's intended behavior.

However, it is worth considering whether a future version of the model could offer the option to reduce the number of warnings, even at the expense of line efficiency.

This would relieve the production manager from having to manually adjust worker allocations, especially in cases where there is ample buffer time before order deadlines, making such adjustments operationally unnecessary.

During the validation sessions focused on production planning, several qualitative comments were collected from the users involved. Although the demonstrator is still at a relatively low TRL, the feedback provided valuable insights into the system's usability, clarity, and decision logic. The following representative comments reflect a constructive attitude and highlight areas for potential improvement in future iterations of the tool.

"The Gantt chart provides detailed information, but I initially found it difficult to understand why certain tasks were split into multiple segments. Including a legend or contextual tooltip could improve clarity.": Refers to the fragmentation of tasks caused by shift changes or worker reassignments.

"I expected the system to maintain operator assignments within the same shift. Introducing a constraint or option to preserve shift consistency would enhance realism.": Refers to unnecessary changes in operator assignments during a single shift.

"In some cases, the schedule includes unexplained idle periods. A brief explanation or visual indicator of the cause would help users interpret these gaps.": Refers to unjustified pauses in the schedule (e.g., week 36 and 40).

"The concept of 'tardiness' is useful, but the current implementation schedules production to start on the due date rather than finish by it. Clarifying this logic would help align expectations.": Refers to the operational definition of tardiness in the CP-tardiness method.

"The system offers a wide range of configuration options, which is appreciated. However, a comparative visualization of the different allocation strategies would support faster and more informed decision-making.": Refers to the comparison between CP-balanced, CP-tardiness, and CP-makespan methods.

The next section addresses the analysis of resource allocation; the results of the analysis and the conclusions will be discussed later in this document.

2.3.2.7   Resource Allocation analysis

Resource allocation was only briefly addressed in the previous section, where the focus was primarily on the number of workers assigned to each task. In this section, we shift our attention to the selection of specific workers for each activity.

In traditional production settings, worker assignments are typically based solely on medical constraints and ergonomic indices. However, within the FAIRWork project, the allocation process adopts a more comprehensive approach: the selection of workers considers multiple additional factors, including experience, personal preferences, and resilience. This enables the generation of more balanced, sustainable, and potentially more effective allocation plans, both from a productivity and human-centric perspective.

As an example, a detailed analysis of week 36 is provided below, starting from September 4th with a time horizon of 7 days and the "Availability Type" set to "Programmed". The order allocation methods is "Constraint Programming - balanced", while the worker allocation method in this case is "Constraint Programming" in all its versions: Balanced, Experience, Preference, Resilience.

When the worker allocation method is set to Constraint Programming, the system uses a CP solver to assign personnel to production lines. This approach applies constraint-based logic to optimize assignments according to predefined criteria.

Specifically, the Constraint Programming method supports four allocation strategies, each with a distinct objective:

- Balanced: Equally balances three factors — experience, preferences, and resilience — with a 1:1:1 weighting. It aims to ensure that experienced workers are assigned to suitable roles while also respecting personal preferences and adaptability.
- Experience: Prioritizes experience by assigning workers to tasks where they are most skilled, thereby improving productivity and quality.
- Preference: Maximizes worker satisfaction by taking into account their preferences for specific tasks or production lines.
- Resilience: Favors resilience, meaning the ability of workers to adapt to changes or unforeseen conditions.

The result is summarized in the image below:
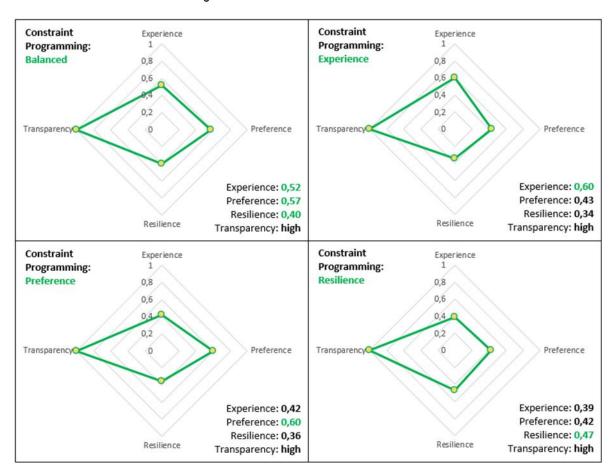
Figure 36 Worker allocation method: "Constraint Programming"

As clearly demonstrated, the four variants lead to different outcomes, since the selection of workers is performed by maximizing the specific objective associated with each method. It is worth noting that in this case, Transparency is high.

Transparency is defined as the degree to which the logic and functioning of a service can be understood by the user. It is evaluated on two levels:

- Global transparency, which refers to the overall system and depends on the specific use case.
- Local transparency, which describes how the service works and how its results are derived.

A service based on analytical expressions and deterministic methods is considered more transparent than one relying on statistical models. Additional factors, such as information about the service's authors and reviewers, also contribute to the overall transparency score.

Let us now move on to analyzing the Worker Allocation Method: Reinforcement Learning, which uses a reinforcement learning approach to assign workers. This method requires moderate computational resources but is more efficient than other simulation-based methods. It learns optimal allocation strategies by interacting with the environment and receiving feedback, allowing it to improve over time and adapt to complex scenarios.
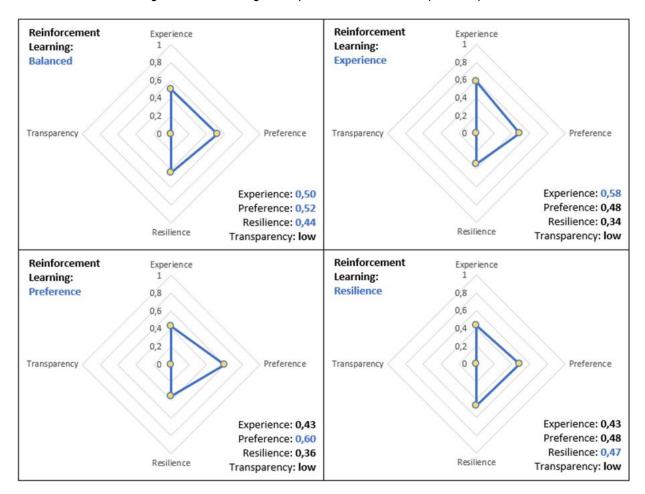
Figure 37 Worker allocation method: "Reinforcement Learning

Also in this case, as the optimization parameter changes, the model responds correctly by maximizing the desired aspect. In this case the Transparency is low.

Finally, we analyze the Worker Allocation Method: Monte Carlo Tree Search, which, as described in the annex, uses a Monte Carlo Tree Search (MCTS) approach to assign workers. This method is computationally intensive and requires more time than the other available methods. It explores possible allocation strategies by simulating different decision paths and evaluating their outcomes, progressively refining its choices to identify the most promising assignments.
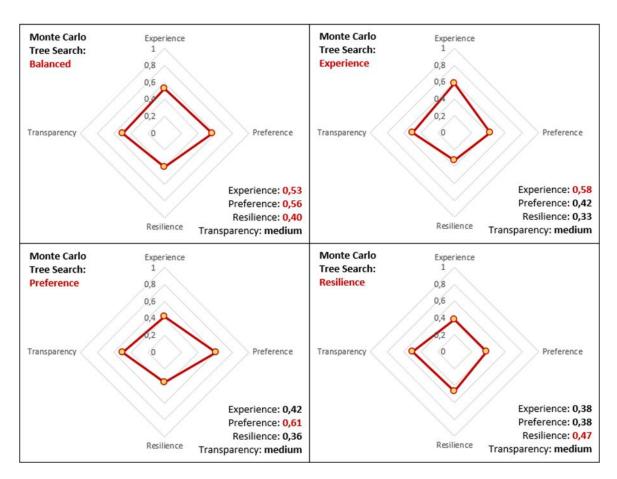
Figure 38 Worker allocation method: "Monte Carlo Tree Search

In this case as well, the model performs excellently. It is important to highlight that while the first two methods (Constraint Programming and Reinforcement Learning) have short execution times, approximately 30 seconds, which is well within the project's requirement of under one minute, the Monte Carlo Tree Search method is significantly more computationally demanding, resulting in a much longer execution time of around 20 minutes.

For this reason, this method is not suitable for real-time planning at the beginning of a shift, especially when considering actual worker attendance. Instead, it is more appropriate for end-of-week planning, where it can be used to prepare the schedule for the upcoming week based on the programmed availability of operators.

For this method, transparency is considered medium.

Finally, for completeness, the image below presents a comparison of the three methods in the case where the preference criterion is set to "balanced", which was identified as the preferred choice by the plant personnel.
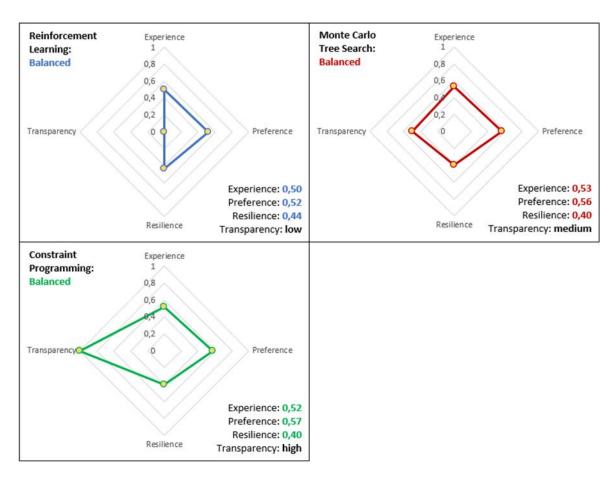
Figure 39 Worker allocation method: "Balanced"

The plant personnel expressed a high level of satisfaction with the results obtained during the validation phase. The proposed approach to resource allocation represents a significant improvement over current practices, which typically do not consider parameters such as workers' experience, preferences, and resilience, mainly because such data are not usually available in a structured format. The integration of these criteria, made possible by the FAIRWork system, was considered both appropriate and preferable, as it introduces a more human-centric and sustainable dimension to operational planning. Regarding the planning outcomes themselves, the results generated by the system were consistently aligned with those achievable through traditional methods, thereby confirming the model's validity. In this respect, the new approach does not replace or undermine existing logic but rather enriches it, offering tangible added value. The tests were conducted over a wide range of weeks, and although it was not possible to identify a universally superior method, all the generated plans were found to be correct and consistent with the plant's operational goals.

During the test sessions, several participants from the plant shared their impressions regarding the new resource allocation approach. The feedback was generally positive and highlighted the potential benefits of integrating human-centric criteria into the planning process. Below are five representative comments:

"It's the first time I've seen a system that considers operators' preferences and resilience. Even though we don't yet have this data structured, the concept is very promising and opens new possibilities."

"Normally, we assign workers based only on availability and medical constraints. This approach is more comprehensive and could improve well-being on the shop floor."

"We tested several allocation strategies and all produced consistent results. It's hard to say which one is best but knowing the system can adapt to different criteria is a major advantage."

"The fact that the system generates valid plans even according to traditional criteria is reassuring we're not losing anything, but we're gaining in quality and sustainability."

"I was impressed by the ability to compare different allocation logics. It would be helpful to have an automatic summary to support the final decision."

The following section presents the results of both analyses.

2.3.2.8 Test results

During the validation phase of the FAIRWork decision-making model at the Stellantis site, tests were conducted over sixteen operational weeks to assess the system's ability to support complex decision-making in a real production environment. For each week, participants assigned a score from 1 to 10 across two key dimensions: the quality of production planning and the effectiveness of workforce allocation. The average results were 8 for planning and 8.5 for allocation, both exceeding the acceptability threshold of 7, thus confirming the robustness of the system.

The methodology followed the protocol defined in Deliverable D5.3, with specific reference to Section 5.3.3, which outlines both subjective evaluation (via user scores and comments) and objective assessment through technical KPIs.

The table below summarizes, week by week, the evaluations provided by the line personnel:

Test	Week (from Monday)	Planning	Allocation	Comment
01	Week 36 (September 4)	7,5	9.5	Good planning overall, but unjustified pauses and excessive task fragmentation were observed.
02	Week 37 (September 11)	8	8.5	Improved performance, though some inefficiencies in worker allocation remain.
03	Week 38 (September 18)	8,5	8.5	Solid planning with minor issues related to worker warnings.
04	Week 39 (September 25)	8,5	7.0	Like Week 38, with good task distribution and few problems.
05	Week 40 (October 2)	7	9.0	Overproduction and unexplained breaks in scheduling affected the score.
06	Week 41 (October 9)	8	9.0	Strong planning, but frequent worker shortage warnings.
07	Week 42 (October 16)	8,5	7.5	Effective planning despite some resource allocation issues.
08	Week 43 (October 23)	8	7.0	Efficient scheduling with some minor inefficiencies.
09	Week 44 (October 30)	6,5	8.5	Significant overproduction and the highest rate of worker shortage warnings.
10	Week 45 (November 6)	7,5	8.5	Planning was too early compared to due dates, but manageable.
11	Week 46 (November 13)	9	7.0	Excellent week with minimal warnings and well-balanced planning.
12	Week 47 (November 20)	8,5	9.0	Good performance with a few manageable warnings.
13	Week 48 (November 27)	8	9.5	Effective planning with some allocation issues.
14	Week 49 (December 4)	8	9.5	Good distribution, though some pauses and warnings were noted.
15	Week 50 (December 11)	8,5	9.5	Robust and well-balanced planning.
16	Week 51 (December 18)	8,5	8.0	Final week was well-managed with few reported issues.
	Average result:	**8**	**8.5**	

Figure 40 Decision-Making Model evaluation, results

The main indicators monitored were:

- Decision Time: average time to generate a plan was under 1 minute for CP and RL methods, meeting the operational target.
- Alternative Plans: the system successfully provided up to 5 alternative solutions when production was feasible with available personnel, as required by the protocol.
- Planning Score: average of 8/10, above the target of 7.
- Allocation Score: average of 8.5/10, also above the target of 7.

In addition to the short-term indicators evaluated during the test sessions, two medium-to-long term KPIs have been identified to assess the broader impact of the FAIRWork system. These indicators will not be immediately measurable within the project timeframe but will require post-project monitoring based on end-of-period operational data.

- Unplanned Absenteeism [%]: This KPI is directly linked to worker well-being and reflects the percentage of person-hours during which employees were unexpectedly absent from work. The calculation method is detailed in Deliverable 2.1 and includes both paid and unpaid absences. The acceptable target is no worsening compared to the current baseline, with an optimal goal of a 2.5% reduction. This indicator is particularly relevant to the "Workload Balance" scenario, where improved task distribution and consideration of worker preferences and resilience are expected to reduce stress and fatigue.
- Energy Consumption [kWh/Ton]: This KPI measures the total energy consumed per ton of components produced, with data collected from the press lines. The methodology is also described in Deliverable 2.1. The desired outcome is a 5% reduction in energy consumption, which would indicate more efficient production planning and better utilization of resources. This KPI is closely associated with the "Production Planning" scenario, where optimized scheduling can reduce idle times, unnecessary machine setups, and energy waste.

These indicators will serve as a reference for evaluating the long-term sustainability and operational benefits of the FAIRWork system beyond the scope of the current validation activities.

2.3.3 Cost-Benefit Analysis

The validation of the DAI-DSS system at CRF (Stellantis – Mirafiori) revealed a significant potential for operational improvement, both in production planning and workforce allocation. Although consolidated economic data is not yet available, preliminary observations indicate measurable benefits in terms of time savings, resource optimization, and enhanced decision-making efficiency.

Estimated Benefits

- **Planning time reduction**: the average time required to generate a weekly plan decreased from approximately 2 hours (manual) to less than 1 minute (DAI-DSS), with an estimated monthly saving of 8–10 hours.
- **Improved resource allocation**: the integration of criteria such as experience, preferences, and resilience led to a more balanced distribution of operators, with potential positive effects on productivity and well-being.
- **Decision quality**: with an average score of 8.5/10 in workforce allocation, the system supports more informed and transparent decisions, reducing the need for manual adjustments.

Potential Economic Impact (Estimates)

- **Energy optimization:** over the past two years, due to the automotive sector crisis, production volumes have been irregular and variable. In a reference period, energy consumption on press lines resulted in significant operational costs. Applying the 5% reduction target outlined in Deliverable D2.1, the potential saving in that context was estimated between €10,000 and €15,000.
- **Absenteeism reduction:** the adoption of human-centric criteria may reduce unplanned absenteeism by up to 2.5%, improving workforce stability and lowering costs related to overtime and reassignments.
- **Time savings:** assuming 10 hours/month saved in planning and allocation, with an average cost of €40/hour (technical staff), this benefit can be scaled across multiple plants achieving a high impact.

2.3.4 Summary of Findings

2.3.4.1 Human Machine Interface (HMI) Validation

The validation of the FAIRWork system's HMI with DAI-DSS fully achieved its intended objectives, demonstrating high effectiveness in both operational usability and perceived user satisfaction. The quantitative test, with an overall average of 89%, exceeded the minimum target of 80%, confirming that users were able to successfully complete many of the assigned tasks. The qualitative test, based on the SUS standard, recorded an average score of 84%, well above the 73% acceptability threshold, corresponding to an "excellent" rating. These results indicate that the interface is mature and ready for adoption in real production environments. However, the analysis also highlighted some areas for improvement, particularly in parameter configuration, algorithm selection, and result interpretation. User feedback suggests the introduction of support tools such as interactive tutorials, contextual descriptions, and improved consistency across modules.

In summary, the HMI proved to be robust, intuitive, and well-received, with optimization opportunities focused on enhancing the user experience and the transparency of AI-driven decisions.

Based on the results, the following recommendations are proposed:

- Introduce an onboarding module for new users, including guided tutorials and contextual tips.
- Make the logic behind allocation algorithms more explicit, by providing brief descriptions next to each available option.
- Improve visual and functional consistency across different interface sections to ensure a smoother user experience.
- Enhance the readability of results and alerts, with clearer visualizations and more prominent messages.
- Provide an "assisted mode" for less experienced users, offering step-by-step guidance for more complex operations.

If implemented, these improvements could further strengthen the effectiveness of the HMI and support its broader and more informed adoption in industrial settings.

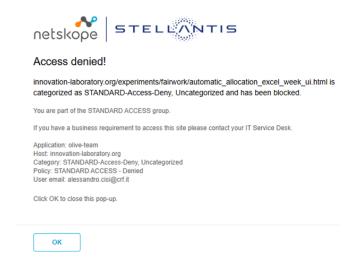Intermittent error that needs to be investigated:

Figure 41 Intermittent error

2.3.4.2 Decision-making model validation

The validation of the decision-making model integrated into the FAIRWork system confirmed its effectiveness in supporting complex planning and resource allocation tasks in real industrial environments. The tests, conducted over sixteen operational weeks at the Stellantis plant, demonstrated high-quality outcomes in both production planning (average score: 8/10) and workforce allocation (average score: 8.5/10), both well above the acceptability threshold of 7.

The analysis highlighted several strengths:

- The system's ability to generate alternative plans quickly (under one minute for CP and RL methods).
- Consistency and correctness of the proposed solutions, even when compared to traditional planning criteria.
- Introduction of innovative criteria for workforce allocation (experience, preferences, resilience), which were appreciated by operators for their more human-centric and sustainable approach.

However, some critical issues were also identified:

- Unnecessary task fragmentation and unexplained idle periods in the schedule.
- Misinterpretation of the "tardiness" concept in certain configurations.
- Production scheduled significantly earlier than due dates, potentially overloading storage areas.
- Frequent "too few workers assigned" warnings, requiring manual intervention.

Based on these findings, the following recommendations are proposed:

1. Refine the planning logic, particularly to avoid excessive anticipation of production and to improve the handling of idle periods.
2. Correct the implementation of the "tardiness" method, ensuring that production is completed by the due date, not started on it.
3. Introduce optional constraints to maintain shift consistency, avoiding operator changes within the same shift.
4. Improve system transparency, for example by adding explanatory tooltips in Gantt charts and visual comparisons between allocation strategies.
5. Provide an assisted planning mode, to support less experienced users in managing complex configurations.
6. Consider, in future versions, a feature that suggests reducing the number of shifts (e.g., from three to two) when the system detects significant anticipation in production relative to due dates. While this functionality could not be implemented in the current version—since workers were pre-assigned across three shifts—it represents a promising opportunity to optimize resource usage and reduce operational costs.

In conclusion, the FAIRWork decision-making model has proven to be mature and ready for industrial deployment, with clearly identified areas for improvement. The integration of human-centric criteria adds significant value, aligning with broader goals of sustainability and workplace well-being.

3 EVALUATION OF DAI-DSS APPROACH

3.1 Research Design and Methodological Reflection

To address the specific form and the multifaceted challenges involved in the socio-technical practice of "doing democracy" within companies and to understand the requirements for implementing a DSS within a company, a case study approach was adopted, which facilitated an in-depth analysis of the human factors in the project (Yin, 2014). The case studies enabled a comprehensive investigation of our use cases and explored the potential demand for the democratic design of the decision-support tool. For this purpose, a three-step procedure (see Figure 42) was implemented, which includes steps of onboarding, precision and contextualization. In the Onboarding Phase, the focus was on an on-site visit to the lead case company FLEX, located in Althofen, Austria, to set the foundation of the empirical case study, gain an overview, and collect primary data for initial analysis. The case study, which was carried out in October 2023, focused on a detailed exploration of the use case involving document analysis, on-site observation, and worker interviews. To gain deeper insights into the human perspective and to explore additional options for supporting the democratic implementation of the DSS, the Precision Phase was conducted in July 2024. This phase once again centred on FLEX, our lead case, to ensure alignment and address potential language barriers. The objective was to deepen our understanding of the company and investigate the dynamics of democratization in industry through MAS. To this end, a comprehensive case study was planned, including on-site observation, worker interviews, and a workshop. The final step in our empirical research was the Contextualization Phase, where we aimed to assess whether the findings were systematic and to determine their feasibility for implementation within the targeted companies. This step, conducted in two phases, focused on validating the results from the previous stages to understand how they could support both industry and service partners. The first phase was built around a workshop (fall 2024), which explored the alignment of the findings with the partners designed services, which will be presented in Section 3.3. Since the deployment of different prototypes by the industry partners required time, the second phase was finally carried out to capture recent updates from the implementation of the DSS (done with interviews and reflection units; summer 2025).

Figure 42 Extensive research plan for the investigation of human aspects in AI-guided decision-making in FAIRWork.

3.2 Interdisciplinary Integration and Tool Interplay

Achieving a specifically for companies shaped democratic AI decision-support system is requiring a close integration of social and technical approaches, as the complexities of democratizing decision-making in hierarchical structures had to be investigated through fieldwork in real company contexts. Furthermore, applying digital technologies like MAS to enable more participatory decision-making in companies led to new socio-technical arrangements, which created both opportunities and limitations for enacting democratic features; particularly in how

worker's representation and status are shaped. Thus, the use of AI, such as MAS, to support democratic decision-making emerged as a key focus of this project; its potential could be better understood through the analysis of the case studies.

Fundamentally, democratic decision-making can be seen as an effort to involve mostly all individuals within a group in the decision-making processes and to presumably avoid illegitimate centralization and concentration of power in these. At the same time, questions of participation and representation became increasingly decisive and were addressed in the subsequently designed method (Figure 43). The AI must be interpreted as a technical form of representation. On the one hand, this offers new opportunities of representation as the AI can be supported with a broad range of parameters representing the present status of workers and production lines. On the other hand, this might go along with specific shortcomings of representativeness as the quality of the technical representation depends on the quality of the parameters set up so far. Moreover, although the quest of representation might be answered sufficiently on the technical level, there might be problems of legitimacy emerging in the respective socio-organizational setting. Thus, a thorough analysis of the socio-technical setting is key. Thereby, the crucial question of representation must be set up as part of a socio-technical practice and must be part of the socio-technical modelling with and through AI. Given this context, the specific method presented here is a reflexive method of analysing the embedding of the AI tool into the respective socio-technical settings.
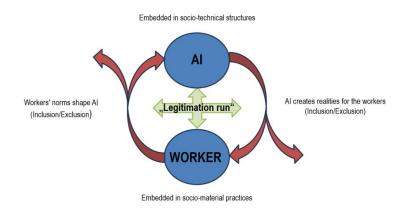

Figure 43 Democratic Exploration Space: The Question of Relationality

Developing a democratic decision-making tool for balancing the tension between social aims of representation and the technical form of representation is decisive. Doing so, it is recognized that shop-floor workers are embedded in the socio-material practices of their specific work environment. Consequently, a distinct necessity emerged for tailored representation to address their unique needs. Representation within this context may take shape through two primary channels: either workers autonomously articulate their needs, or experts are tasked with identifying and expressing these requirements (Figure 43). In the various steps of the case study research, workers' needs have been expressed both directly by themselves and indirectly through their workers' council and managers, offering a comprehensive overview in this regard.

Given the potential variability in worker fitness and alertness, digital tools may offer advantages over workers' self-assessment in certain situations, while personal experiences may provide deeper insights in others. Our approach integrates both methods, employing digital tools alongside workers' self-assessment to comprehensively address the organizational issues. This integration extends to the operation of digital tools relying on specific inputs such as parameters, factors, and indicators representing the workers. These elements are integral components of the AI representation, crucial for enhancing its functionality and informing decision-making processes. AI creates specific realities for workers by making decisions using its internal support system. This raises questions about the reliability of outcomes for the workers and whether they are adequately represented by the AI, referred to as MAS in our case. Consequently, tensions emerge while confronting decisions made by the system with the expectations of the

ones being represented by the system. To understand how these tensions are resolved, consideration of the concept of legitimation is useful. This involves examining the extent to which social processes and their monitoring are valued, as well as the opportunities workers must represent themselves differently. Thus, it is imperative to focus on the democratic character of the system and the legitimation process, which seeks to address the tension between social demands for representation and technical forms of representation.

The findings indicate that democratizing decision-making in socio-technical settings can entail the integration of essential democratic features in the implementation process. This ensures that the implementation of the DSS not only enhances efficiency but also improves the company and empowers employees to engage in meaningful participation. One of our key results in conducting the case studies consists in identifying additional factors which have been regarded as democratic features in the respective cases (Figure 44). Besides the primary features, Learning Implementation is recognized as a key factor that fosters democracy because it enables a long-term, comprehensive learning period from scratch, in which all employees are actively involved. Such implementation is less likely to be denied or forgotten, and any limitations of the digital tool can be identified quickly and resolved. Customizability also enables last-minute adjustments for customers and decision-makers, giving them more control over how the system operates and enhancing the system's flexibility. Clear responsibilities (or rather accountabilities) are always a relevant factor in the development of organizations. This is even more the case when organizations are undergoing a fundamental change, like in the case of such an implementation of a digital technology for supporting decision-making. Additionally, accountability is an important democratic feature because it fosters transparency and fairness. This allows individuals to track the decisions made by the system and ensures that decisions can be changed or improved if necessary.
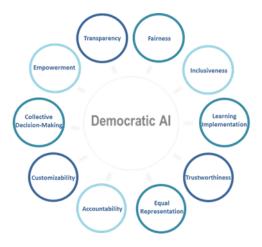

Figure 44 Relevant Features of democratic AI in companies.

Consequently, these features empower employees by providing tools and opportunities for active engagement in decision-making, ensuring their voices are heard and fostering a more inclusive, collaborative environment that promotes collective decision-making. At the same time, the dimensions of relevant features of democratic AI in companies as shown in Figure 44 make it clear that a multitude of relevant criteria must be considered in processes of democratization in companies. However, the criteria do not behave consistently in a linear fashion (in the sense that achieving one would simultaneously imply improving the other criterion). Rather, the situation is much more complicated. In some cases, these criteria are competing to each other. Therefore, two strategies are important. Firstly, these criteria must be considered in the MAS, and the procedure must be supported. Secondly, the social quality of the implementation processes of such tools plays a central role. Otherwise, this process would come into sharp contradiction to the articulated relevant criteria for the evaluation of democratization

3.3 Peer Review and Cross-Evaluation of Methods

After completing two steps of our case study, it was imperative to have our findings assessed with the project partners in order to evaluate the extent to which the results are aligning with the criteria used in their focus area, and to examine the quality and credibility of the outcomes. This evaluation process helped reveal the limitations and blind spots of the applied methods and contributed to their refinement based on the partners' expertise. In particular, the features of democratic AI mentioned above ultimately identify the central objectives of a model for democratic decision-making in companies. Nevertheless, the key challenge remains in how these features can be operationalized in concrete terms. This process of operationalization is highly context-dependent, varying with each company's specific application. To address this, the contextualization step was conducted in two phases, further advancing the project objectives.

3.3.1 Interactional method

Since the critical question was how the DSS can be democratic and how to operationalize the features for democratizing companies through digital tools, a corresponding workshop was held in the first phase, during which partners actively participated in providing answers. In November 2024, an online workshop was held via a virtual collaboration board and included two tasks:

- The first task asked collaborators to match the democratic AI features (identified through the case study) with their services or research and to explain how each feature could be implemented (Figure 45).
- The second task listed the expectations gathered from the workers and asked our colleagues to indicate which of these could be addressed through their ongoing work (Figure 46).

Regarding the first task, the features of democratic AI in the DSS can be applied differently, as the developed services serve various purposes. For example, Inclusiveness may refer to accessibility for all workers in one service, while in another, it may emphasize gender inclusion, or it may not be addressed at all if the service is designed primarily for experts. Notably, features such as transparency, trustworthiness, and fairness are strongly considered across different services, for instance using explainable AI, anonymized data, and trackable decision-making processes. The accountability of services is also perceived in different ways; for instance, it may be assessed through feedback from end users, the use of signatures, or the identification of responsible parties. The ability to adjust and adapt within the designed service is also significant, as it can be achieved through customization for different production lines within the same use case, modifications to the model structure, and the flexibility of the use case itself. However, features such as collective decision-making and equal representation have been not strongly addressed, which may be due to the different purposes these AI systems are designed for. Moreover, since such systems are primarily intended to answer users' questions, offer options, and function as assistants, they may struggle to fully support all the identified features of democratic AI. Ultimately, the feature of Learning Implementation, identified as a novel aspect through the empirical analysis, was not given significant attention by the partners. This is not surprising, as the feature emerged only during the empirical research, and the digital tools in question are self-learning, requiring minimal training or familiarization. However, this feature could be relevant for other training processes that are more time-consuming and complex.
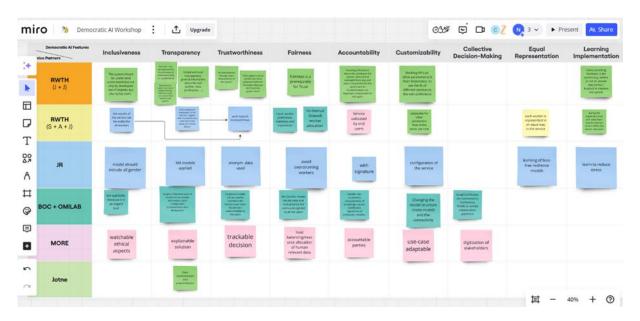
Figure 45 Workshop Board Activity - Task 1: Evaluating Democratic AI Features Across Partners' Focus Areas

Expanding on the operationalization of findings through the workshop process, we asked our project partners, within the second task, to identify which expectations mentioned in both the Onboarding and Precession phases are applicable to their services and research. As shown in Figure 46, simplifying system usage, empowering staff for decision-making, reducing human intervention, and fostering continuous learning during tool deployment were the most frequently mentioned. Collaborative decision-making with AI, indicating resource availability and locations, ensuring fair treatment of employees, balancing team contributions, and enabling data and workflow review have been also highlighted.
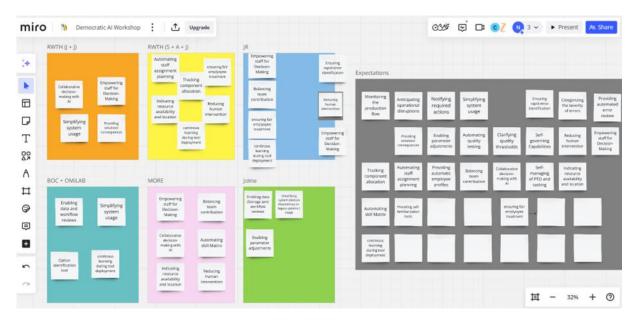
Figure 46 Workshop Board Activity - Task 2: Evaluating the Fulfilment of Workers' Expectations Across Partners' Focus Areas

3.3.2 Interview method

The second phase of the contextualization process involved expert interviews and reflexive units for both the development and implementation of the use case AI Maintenance Support. One important aspect hereby was that

given how one of the companies manages approximately 8,000 unplanned repair calls annually, primarily involving complex machinery and processes, maintenance tasks are often carried out by relatively inexperienced personnel. The objective of the use case was to develop an AI-based tool designed to support these workers in maintenance activities, thereby aiming to reduce both repair time and operational downtime. After the development and activation of the first prototype for preliminary evaluation, feedback was collected from project partners, which informed the further development of the AI tool in a second prototype. The initial evaluation of the second prototype focused on both the written and speech-based modules, with defined performance targets: 100% identical answers for identical questions, 80% correct answers for freely chosen questions on the same topic, and 60% correct answers for freely chosen problem descriptions. According to the expert, five employees including one team leader and four maintenance workers are currently testing the developed AI tool. The primary function of the AI tool is to assist during breakdowns, machine malfunctions, and process errors. It aims to facilitate faster problem-solving by providing direct access to relevant information, thereby eliminating the need to manually search through technical documents. The tool required no special familiarization period, as it is designed to be self-explanatory and easy to use, particularly for inexperienced workers. However, since the AI tool is intended to respond to workers' questions, it is crucial how the questions are formulated; this is the aspect that still requires some learning. While the tool generates responses that workers can follow, these responses can be verified against the downloadable user manual.

The testing outcomes revealed that, while the speech module tests were unsuccessful due to high noise levels in the production hall, the written tests produced considerably better results. Among the different test types, the system performed best with identical questions, delivering 90% of the expected responses. However, challenges remain in handling freely chosen questions on the same topic, with only 64% of answers being accurate. To address these issues, the evaluation process, including an assessment of the system's basic knowledge base, is planned to be repeated and reverified. The third test is still in progress.

The AI maintenance support functions as a question-answering system, effectively guiding users toward more detailed information on how to handle breakdowns. It provides step-by-step procedures outlining what actions should be taken in response to specific situations and errors. From this perspective, the transparency of the system's answers is considered acceptable. Users can decide on and directly implement the tool's responses if they are plausible and relevant, as the tool is designed to be reliable, and its suggestions are generally followed without hesitation. This reflects the perceived trustworthiness of the tool in supporting decision-making processes. In cases of uncertainty or difficulties that the system is unable to support, users are expected to consult with their supervisors to ensure accurate decision-making. Although the AI responses do not fully meet expectations, a key factor hindering the system's ability to provide adequate support is poorly formulated questions. Therefore, it is important to reformulate and present the system with alternative and precise queries, and to further verify the underlying causes.

Working with the AI tool in FLEX is comparable to retrieving information from an instruction manual, as it provides step-by-step guidance for inexperienced workers in handling specific machine issues and resolving errors. However, the AI tool enables much faster access to this information than manually searching through technical documentation or handbooks. Since the responses are drawn from verified knowledge bases that have been specifically curated for integration into the tool, it supports greater accountability and helps ensure that the answers provided are responsible and based on validated information. Furthermore, the broader implementation of AI maintenance support across the entire maintenance area, beyond its current training-focused purposes, depends on cost considerations and the acceptance of experienced personnel. While the system facilitates faster and more efficient familiarization for new workers, enabling more accurate decision-making, its utility for experienced employees may be limited unless it is further developed to incorporate advanced features and functionalities.

3.4 Reflection and Lessons Learned

To gather the lessons learned and gain the opinion of the whole consortium, multiple workshops were organized and held with all partners of the FAIRWork Consortium. The following section will display these lessons learned, structured by various parts of the project.

3.4.1 Project Structure and WP Distribution

The structure of FAIRWork was oriented at the Design Science Framework proposed by Hevner et al. (2004). E.g., the Environment was provided through the Use Case Partner and WP5, Research, as well as the Knowledge base were covered through WP3, while the Development and Evaluation were in WP4 and 6, respectively.
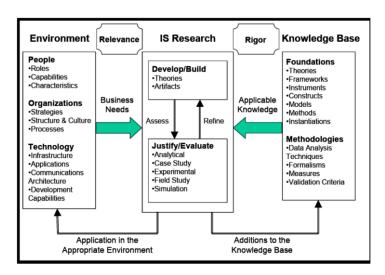
Figure 47 Design science framework by Hevner et al. (2004)

The structure and coordination of the work packages were considered very effective, with clear roles and good internal collaboration among teams. Technical integration and communication between partners supported this, though some interfaces – particularly with WP7 and WP8. Integration across levels (research, prototype, and industry) and work packages was effective, with interfaces and exchange platforms established efficiently. Some challenges existed at the project's beginning, in the understanding of company-specific contexts, suggesting that earlier engagement with industrial partners might have improved alignment. This was solved by targeted alignment meetings and discussions to strengthen inter-partner and inter-disciplinary understanding. Communication of results beyond the consortium can be a challenging area as well; however, in the FAIRWork project, this was approached through multiple dissemination activities detailed in Deliverables D7.1 and D7.2, such as webinars, the innovation shop, industrial and scientific publications.

3.4.2 Environment Approach, Workshops, and Use Cases

In the project, an environment approach was established through iterative and participatory workshops. The use of iterative workshops, onsite visits, and decision modeling supported flexible and dynamic scenario development. Use cases evolved in response to real-world changes, and most participants did not feel constrained by a fixed scenario set. Tools like design thinking and decision models were viewed as effective for identifying priorities and structuring stakeholder input. While the variety of use cases added value, it also introduced a challenge in maintaining a balanced focus across them. The approach was grounded in the Use Case context but designed at a level of abstraction that allows adaptation beyond it, indicating good generalizability of the results.

3.4.3 Prototypes and Demonstrators

Throughout the project, an innovation funnel with multiple iterations was realized. This enabled the creation of approximately 25 laboratory experiments, which in turn led to the development of 10 prototypes, seven demonstrators, and ultimately three to four solutions.
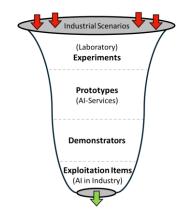

Figure 48 FAIRWork Innovation Funnel

The prototype and demonstrator development process was seen as very valuable and highly rated. Even though the process might seem disorderly initially, with multiple directions pursued before narrowing focus late in the project, this flexibility allowed the team to explore different ideas and prioritize only the most useful ones, leading to well-targeted demonstrators. The approach supported understanding of service usability and deployment needs, and demonstrators proved helpful for communication with external stakeholders. The chosen approach enabled the consideration of a wide field of possibilities, options, and experiments, without unnecessary constrictions. One challenge that arose was deciding when to stop or prioritize prototypes, which required regular discussions and reflections within the consortium, as well as close cooperation and alignment with the use case partners.

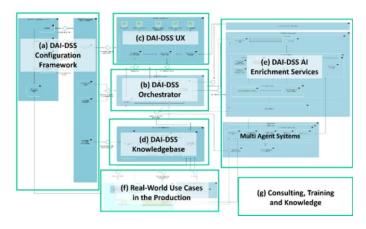### 3.4.4 The FAIRWork Architecture

Figure 49 FAIRWork High-Level Architecture

The FAIRWork architecture was seen as helpful for structuring work and illustrating modularity, especially for guiding activities in WP4. The architecture was developed in multiple iterations (see deliverable D4.1), which allowed for the implementation and integration of further learnings and improvements into the architecture. The high-level architecture (Figure 34) provides an overview of the different building blocks, their connections, and interconnections, making it easier to understand and explain the structure and dependencies. The underlying structure with its distinct components was evaluated positively. All the identified building blocks of the architecture had an integral part in the architecture and during the project. Some blocks, such as the multi-agent subsystem, were noted as especially important, and additional focus on this part would be a future opportunity. The abstract

nature of the architecture was both a strength, for avoiding vendor lock-in, and a limitation when it came to usability in specific contexts. Despite this, the architecture enabled deployment and testing of services at the required maturity level (TRL 6/7), and it appears capable of incorporating general AI developments via enrichment.

3.4.5 Exploitation and Ecosystems

For exploitation, value-adding networks were utilized, and three separate exploitation ecosystems were established referring to "Key Exploitable Results", each under the leadership of a technological customer owner or community partner. The exploitation ecosystems were presented in D8.3 and reflected again in D8.4 to incorporate learnings from workshops and to improved and specify the KERs. In general, the KERs apply a "value adding network approach" in which offerings are joined to create a network that enhances value. The exploitation ecosystem concept was generally well-received and viewed as a clear and structured way to package project outcomes. The strategy of organizing ecosystems around major content/business partners was considered feasible and effective, particularly from a development standpoint. Consortium partners noted that this framing helped clarify the market potential of results and fostered cooperation across EU partners. The approach supported the modular and flexible integration of outputs, and the coverage of project results, especially through Key Exploitable Results (KERs), was deemed adequate. For some, this was a novel concept, adding value by contextualizing technical outcomes within a broader exploitation strategy.

3.4.6 Innovation Shop and Innovation Items

The integration of all levels of the project was carried out regularly through (joint) webinars, partner meetings, deliverables, and explicitly stored in the Innovation Shop. In this way, the results were continuously developed and iteratively integrated on a regular basis. These Innovation Items and the Innovation Shop were regarded as an effective and practical mechanism for representing and communicating project results. Their flexible granularity made them useful across different contexts, and participants appreciated the structured visibility they offered. The flexible nature of the Innovation Shop enabled the integration of new findings and results throughout the project, either expanding existing innovation items, or creating entirely new items. The Innovation Shop was noted as a novel and valuable feature for making outcomes available beyond the project duration. To improve the accessibility of items the initial innovation shop concept was revised and extended with "self-service marketplace" option to buy an instance of the item, if the item is eligible. The items encapsulated a vast range of outputs and helped link research and development efforts. The system facilitated both dissemination and reuse, adding value for future engagement.

4 FUTURE ROADMAP AND LONG-TERM IMPACT

Here, the focus lay on the perception of the project by the Use Case partners, specifically whether and how they can use the project's results, and what effect can be expected from this. Additionally, a roadmap was developed to show the next steps needed to further evolve the Use Cases towards a fully operational paradigm.

4.1 Current Status of the DAI-DSS system at the Use Cases Partners

This subchapter contains a short status of the application of the DAI-DSS at the Use Case Partners, as well as their reflection on the DAI-DSS and its applicability.

Flex Althofen

The DAI-DSS system has been successfully demonstrated at the Flex production site (Althofen), validating one of the main services, "AI Support for Maintenance and Machine breakdown". As mentioned in Chapter 4.3 in greater detail, the use case initially planned for Flex needed to be changed during the project. New use cases were identified for both Flex Althofen and Flex Timisoara; however, with this one, it was challenging to reach the desired extent compared to the original use case.

The system was deemed useful, and the Use-Case Partner plans to implement it into their production environment. However, some challenges remain, for example, input via the speech module is currently not working with the desired reliability. While showing promise, tests of the system by the Use-Case Partner are still ongoing at the time of writing.

The system shows great potential to improve the work in supporting workers at the production site of the Use-Case Partner, which is why it is planned to not only introduce the system into the day to day work and to use it for additional pieces of equipment then currently tested, but to also roll it out to different production sites, where it can help other sites with other knowledge databases.

Flex Timisoara

The DAI-DSS system for Calibration Certification Service has successfully demonstrated its capabilities at the Flex Timisoara production site for two out of four types of Calibration Certificates under study. The aim of the "Verification of Calibration Certificate" service is to enhance and simplify the process of checking and verifying machine calibration certificates. Flex Timisoara perceived the project positively, even though the initial use cases with Robots needed to be replaced, as mentioned in greater detail in chapter 4.3. The new use case, "Verification of Calibration Certificate," was slightly simpler than the one originally planned, but it was chosen since it was possible to demonstrate an immediate impact in that field for the Use Case Partner.

The system has demonstrated its usefulness, and the partner wants to integrate it into their production site. Before a full deployment of the system, however, some adjustments are still necessary: the other two categories of certificate of calibration need to be integrated as well, and the work on the user interface needs to be expanded to make the system easier to use.

When finalized, the Use Case Partner expects a direct effect using the system. Utilizing the system for certificate verification will eliminate human errors and save time and manpower. With some adaptations, the system can also be replicated and expanded to other documents across various industries, which Flex wishes to do, to extend the system to multiple production sites.

CRF

The DAI-DSS system has been successfully demonstrated at the CRF production site (Stellantis – Mirafiori), validating two main services: Production Planning & Resource Allocation. Thus, the project was received very positively by CRF, and the Use-Case Partner remains with concrete interest and plans for the operational adoption of the developed system.

The system and its results were tested in a real production environment, and the evaluation of the Use-Case Partner indicates that they are ready for integration into operational processes. The HMI interface received high ratings in terms of usability (average SUS: 84%) and functionality (quantitative test: 89%). The decision-making model received positive feedback from operators, with average scores of 8 (planning) and 8.5 (allocation).

The usage of the DAI-DSS system achieves a direct effect. Usage of the system improves the resource planning and resource allocation. The effect was direct and received positive feedback from operators. The effects of the system can be leveraged across Europe. The system's modular architecture makes it replicable in other industrial sites across Europe, increasing its potential and usefulness for CRF.

4.2 Next steps and roadmaps for the Use Cases Partners

Mentioned below is a roadmap for the planned operational adoption of the DAI-DSS at the Use Cases, as well as the prerequisites and next steps for each of the Use Case Partners.
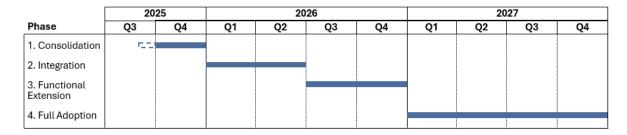

Figure 50 Roadmap for the Operational Adoption of DAI-DSS in FAIRWork Use Cases

4.2.1 Support Access to Information for Machine Maintenance (FLEX Althofen)

The consolidation in Q4 2025 will include bug fixes (e.g., the speech module), user training, and the extension of the knowledge database. The integration phase consists of the implementation in FLEX Net and validating the system in extended production environments. During the functional extension in the second half of 2026, the application of the AI support to other areas is planned. Currently, the system focuses on SMT. The Adoption phase in 2027 will contain the daily use of the DAI-DSS to support maintenance activities. Additionally, a roll out to other product locations in FLEX for implementation is scheduled during this phase.

Enabling Conditions
- Acceptance of AI support in the relevant departments
- Proven improvement of KPIs
- Expansion of the knowledge database to other technical areas
- Acceptable running costs e.g., licenses
- Active involvement of operators and maintenance department.
- Presence of a technical team for maintenance and updates of tool & database
- Definition of baselines and targets for KPIs (e.g., number of maintenance tickets, time reduction for troubleshooting).

4.2.2 Support Validation of Calibration Documents (FLEX Timisoara)

The first phase (consolidation) will include bug fixes and fine-tuning, as well as validation for the third type of calibration certificate. Additionally, the cleanup of the image for the calibration certificate DB is planned for this phase during Q4 2025. The integration will happen during the first half of 2026 and include the integration of the calibration certificate DB within the DAI-DSS. Further development work on the human-machine interface is also scheduled for this period, along with additional validation. The functional extension in Q3 and Q4 2026 will include the application of the DAI-DSS for the fourth type of certificate, and additionally for unstructured calibration certificates. Lastly, full adoption is planned for 2027, with daily use of the DAI-DSS for all types of calibration certificates, and collaboration with other FLEX production sites for the rollout and deployment of the system to these sites.

Enabling Conditions

- Extension of DAI-DSS capability to 3rd type of Calibration Certificate
- Resources availability to continue studies/trials for next types of Calibration Certificate
- Identify a local IT company for integration of DAI-DSS service with Calibration Certificate Data Base and develop the HMI
- Acceptance of cost to continue development
- Active involvement of Calibration Lab in the next steps.
- Availability of JR to sustain the project after FAIRWork is completed
- Definition of baselines and targets for KPIs (successful rate or false calls).

4.2.3 Assist Decisions about Production Planning (CRF)

During the consolidation phase, which CRF plans to initiate in Q3 2025, Bug fixes are planned, as well as improvements to the human-machine interface and transparency through tooltips, contextual explanations, and comparative visualizations. Additionally, user training will take place and introduce interactive tutorials and assisted mode for less experienced users. The integration phase in the first half of 2026 will contain a connection with legacy systems (MES, HR), data flow automation, and validation in extended production environments. The functional extension will happen in the second half of 2026. It will introduce new metrics to the system (e.g., energy consumption KPIs), as well as additional advanced scenarios (e.g., multi-shift simulation). The full adoption of the system is planned for 2027 and assumes daily use of DAI-DSS for planning and allocation. This phase also includes the monitoring of impacts on productivity, well-being, and sustainability.

Enabling Conditions

- Availability of structured data on skills, preferences, and resilience.

- Active involvement of operators and line managers.

- Presence of a technical team for maintenance and updates.

- Definition of baselines and targets for KPIs (e.g., absenteeism, energy consumption).

5 SUMMARY AND CONCLUSIONS

Summarizes the main results of the demonstration, highlights the observed benefits and emerging issues. Provides recommendations for the future adoption of DAI-DSS in real production contexts and suggests possible future developments of the system.

5.1 Improve Information Access to Support Maintenance (FLEX Althofen)

Summary of results:

The evaluation of "Improve Information Access to Support Maintenance" use case shows positive results across the diverse levels. In level 1, standardizes questions are used the test phase achieved a 100% success rate. Responses were consistent and reproducible across multiple independent runs. The system demonstrated high reliability and stability when operating under structured conditions. This confirms its robustness in handling well-defined input scenarios. In level 2, Topic-Based Free Questions are used to assess the system's ability to handle freely formulated but thematically focused questions, simulating more natural user interaction. The system's performance of the first prototype was below expectations. In the optimized trial of level 2 using the extended prototype, including an enriched the knowledge base with relevant domain-specific content and training sessions for test participants to improve question formulation, the system's performance exceeded predefined benchmarks. Responses were accurate, relevant, and context aware. In level 3, the idea was to evaluate the system's flexibility, creativity, and problem-solving capabilities when confronted with open-ended, user-defined challenges, particularly within the SMT (Surface-Mount Technology) production domain. However, responses were inadequate and lacked sufficient depth in SMT-specific content, restricting its ability to generate meaningful of the first prototype. The improved prototype (same as in level 2) with extended knowledgebase and domain-specific questions the system's answers improved. This highlights the necessity that 1) the input must be well-defined and domain-specific and 2) underlying knowledge infrastructure must be sufficiently developed when using the AI.

Due to the inability to isolate clean audio input and the non-deterministic nature of acoustic interference, the speech module could not be reliably assessed. As a result, the evaluation scope was narrowed to focus exclusively on text-based input modalities (keyboard, touchscreen), which provided controlled and reproducible data for analysis.

Across all levels, the system-maintained response times under 10 seconds, meeting real-time production requirements meeting the performance criteria.

In general, using the AI prototype improved the understanding of AI and enabled employees to deepen their knowledge on how to use AI, supporting FLEX in the Introduction of AI into the company.

Main Benefits:

The evaluation demonstrates that the AI system is highly capable under the right conditions. Key strengths include:

- Robust performance with structured input

- Adaptability through iterative learning and database enhancement

- Strong domain-specific problem-solving capabilities

- Reliable response times suitable for real-time interaction

Critical Issues:

- Gaps in the knowledge database related to specific production topics.

- Insufficient user training, leading to inconsistent or poorly formulated questions.

Recommendations for Future Adoption:

- Expand the knowledge base with deeper SMT-specific content and adjacent domains (e.g., AOI, reflow soldering, maintenance workflows).

- Implement structured user training programs to ensure effective system interaction and question formulation.

- Future development of the Machine maintenance prototype should explore advanced speech recognition technologies, such as directional microphones, noise-canceling algorithms, or multimodal input fusion.

- Pilot the system in live production scenarios, focusing on high impact use cases (e.g., fault diagnosis, operator assistance).

- Introduce a feedback loop, allowing users to rate responses and contribute to continuous system improvement.

5.2 Support Validation of Calibration Documents (FLEX Timisoara)

Summary of results

The evaluation of the Calibration Certificate Service for the first two categories of Calibration Certificates has yielded positive results that affirm the system's stability and repeatability, ensuring consistent performance. The service demonstrates a high confidence rate, with less than 5% of false calls necessitating manual human verification.

This automated solution is designed to significantly reduce review time while minimizing human error, ultimately enhancing accuracy and reliability in the certificate review process. By streamlining these operations, the service not only strengthens internal quality assurance but also liberates staff who were previously engaged in manual verification tasks. As a result, employees are empowered to focus on more value-added activities, contributing to overall operational efficiency and effectiveness within the organization.

Main Benefits:

- Significant time savings are achieved through automation.

- The service effectively eliminates human errors associated with manual verification.

Critical Issues:

- Current limitations in the Human-Machine Interface (HMI) hinder ease of use, particularly in selecting Calibration Certificate types and tracking progress during multiple file processing.

- The absence of integration with local SQL databases (CDB) restricts the system's potential effectiveness and operational efficiency.

Recommendations for Future Adoption:

To maximize the benefits and facilitate broader acceptance of the Calibration Certificate Service, the following recommendations are proposed:

- **Enhance the HMI:** Develop a more intuitive interface that includes easy selection options for Calibration Certificate types and incorporates a progress bar to indicate processing completion. This would improve user experience and efficiency.

- **Integrate with Local SQL CDB:** Establish connectivity with local SQL databases to enable seamless data management and facilitate real-time data retrieval for verification processes.

System Development Perspectives:

Future development of the Calibration Certificate Service should focus on refining user experiences and enhancing system capabilities to the last 2 category of calibration certificates. By addressing the issues related to the HMI and database integration, the service can evolve into a more comprehensive solution that not only verifies Calibration Certificates but also supports a broader range of documentation processes across different sectors. Additionally, ongoing feedback from users can guide iterative improvements, ensuring that the system remains aligned with user needs and operational goals.

5.3 Assist Decisions about Production Planning (CRF)

The validation of the decision-making model integrated into the FAIRWork system confirmed its effectiveness in supporting complex planning and resource allocation tasks in industrial environments. The tests, conducted over sixteen operational weeks at the Stellantis plant, demonstrated high-quality outcomes in both production planning (average score: 8/10) and workforce allocation (average score: 8.5/10), both well above the acceptability threshold of seven.

The analysis highlighted several strengths:

- The system's ability to generate alternative plans quickly (under one minute for CP and RL methods).

- Consistency and correctness of the proposed solutions, even when compared to traditional planning criteria.

- Introduction of innovative criteria for workforce allocation (experience, preferences, resilience), which were appreciated by operators for their more human-centric and sustainable approach.

However, some critical issues were also identified:

Unnecessary task fragmentation and unexplained idle periods in the schedule.
Misinterpretation of the "tardiness" concept in certain configurations.
Production scheduled significantly earlier than due dates, potentially overloading storage areas.
- Frequent "too few workers assigned" warnings, requiring manual intervention.

Based on these findings, the following recommendations are proposed:

1. Refine the planning logic, particularly to avoid excessive anticipation of production and to improve the handling of idle periods.
2. Correct the implementation of the "tardiness" method, ensuring that production is completed by the due date, not started on it.
3. Introduce optional constraints to maintain shift consistency, avoiding operator changes within the same shift.

www.fairwork-project.eu   Page 76 of 77

4. Improve system transparency, for example by adding explanatory tooltips in Gantt charts and visual comparisons between allocation strategies.
5. Provide an assisted planning mode, to support less experienced users in managing complex configurations.
6. Consider, in future versions, a feature that suggests reducing the number of shifts (e.g., from three to two) when the system detects significant anticipation in production relative to due dates. While this functionality could not be implemented in the current version—since workers were pre-assigned across three shifts—it represents a promising opportunity to optimize resource usage and reduce operational costs.

In conclusion, the FAIRWork decision-making model has proven to be mature and ready for industrial deployment, with clearly identified areas for improvement. The integration of human-centric criteria adds significant value, aligning with broader goals of sustainability and workplace well-being.