DES HE SESAR ASTAIR exploratory research report

Deliverable ID:	**D5.2**
Project acronym:	**ASTAIR**
Grant:	**101114684**
Call:	**HORIZON-SESAR-2022-DES-ER-01**
Topic:	**HORIZON-SESAR-2022-DES-ER-01-WA1-1**
Consortium coordinator:	**ENAC**
Edition date:	**29 August 2025**
Edition:	**01.01**
Template edition:	**01.00.00**
Status:	**Official**
Classification:	**PU**

Abstract

This document presents the results of the final validation activity, as part of the third validation exercise, carried out in the ASTAIR project. The ASTAIR concept aims to support the automation of airport ground operations, particularly engine-off taxiing, by integrating adaptive algorithms and tools that foster effective human-machine collaboration.

This report details the assessment of the ASTAIR solution through a Human-in-the-Loop Real-Time Simulation (HITL RTS), supported by Fast-Time Simulations (FTS) and a Final Workshop involving

operational stakeholders. The exercise explored the system's usability, operational feasibility, impact on controller workload, and contribution to safety, predictability.

The evaluation examined how human operators interacted with the system in realistic airport scenarios, including non-nominal situations, and considered the clarity of AI-driven decisions, the ability for human intervention, and the alignment with existing procedures.

This report reflects the final step in the ASTAIR exploratory research and confirms whether the concept satisfies the criteria to reach TRL1, building on the foundations set out in the Exploratory Research Plan. The validation outcomes confirm the feasibility of the basic principles, while identifying areas requiring further development to support progression towards higher maturity levels.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Authoring & approval

Author(s) of the document

Organisation name	Date
Deep Blue	27/08/2025
ENAC	13/06/2025
TUD	17/06/2025

Reviewed by

Organisation name	Date
Deep Blue	27/08/2025
ENAC	20/06/2025
TU Delft	17/06/2025
Groupe ADP	16/06/2025

Approved for submission to the SESAR 3 JU by

Organisation name	Date
Deep Blue	27/08/2025
ENAC	20/06/2025
TU Delft	18/06/2025
EUROCONTROL	20/06/2025*
Groupe ADP	20/06/2025

*Silent approval

Rejected by[1]

Organisation name	Date

Document history

Edition	Date	Status	Organisation author	Justification
00.01	04/04/2025	Draft	Deep Blue	Document set up.
00.02	08/06/2025	Draft	Deep Blue	Liability Assessment implemented.

[1] Representatives of the beneficiaries involved in the project

EUROPEAN PARTNERSHIP
Co-funded by the European Union

00.03	09/06/2025	Draft	Deep Blue	Document ready for review by partners.
00.04	16/06/2025	Draft	Deep Blue	Partner's comments implemented.
01.00	20/06/2025	Official	Deep Blue	Document approved and submitted.
01.01	29/08/2025	Official	Deep Blue	Document updated for resubmission and submitted.

Copyright statement

Disclaimer

ASTAIR

AUTO-STEER TAXI AT AIRPORT

ASTAIR

This document is part of a project that has received funding from the SESAR 3 Joint Undertaking under grant agreement No 101114684 under European Union's Horizon Europe research and innovation programme.

Table of contents

EUROPEAN PARTNERSHIP

Co-funded by the European Union

List of figures

EUROPEAN PARTNERSHIP

Co-funded by the European Union

List of tables

1 Executive summary

This Exploratory Research Report (ERR) describes the validation activities and their outcomes, in respect to the third validation exercise. The project progressed from its initial Technology Readiness Level (TRL) 0, through the development of its knowledge base and definition of the technology concept, advancing towards TRL1.

The validation exercises focused on evaluating the ASTAIR solution across the following main Key Performance Areas (KPAs): Human Performance, Environmental Sustainability, Cost Efficiency, Safety, Operational Feasibility, Performance, Acceptability, and Liability. The validation followed an iterative Human-Centred approach, aimed at demonstrating and validating both the algorithmic tool and the Human-Machine Interface (HMI) concept, thereby supporting the solution's development from concept to functional prototype level.

The ASTAIR project investigates the feasibility and implications of introducing automation to airport ground operations, with a particular focus on supporting engine-off taxiing procedures. It explores how adaptive planning algorithms, human–machine interfaces, and collaborative AI can be used to optimise surface movement operations within A-CDM and A-SMGCS environments. The project addresses key Research and Innovation (R&I) needs related to workload reduction, sustainability, and improved predictability, while ensuring that controllers remain in the loop.

The validation approach encompassed three targeted exercises designed to: (1) evaluate the scope and use cases of the ASTAIR solution with the involvement of project stakeholders; (2) assess the design and implementation of the ASTAIR solution through prototype demonstrations; and (3) validate the ASTAIR solution via a Human-in-the-Loop Real-Time Simulation (HITL RTS), Fast-Time Simulations (FTS), and a Final Workshop with stakeholders.

The ASTAIR project validated the concept through three validation exercises:

- Validation exercise #01: involves three workshops where ASTAIR's stakeholders validated the scope and use cases of the project. The exercise outputs the set of use cases and updates the scope for the solution.
- Validation exercise #02: involves two workshops where ASTAIR's end-users evaluated the ASTAIR's solution design and its implementation. The exercise outputs the results to improve the solution's HMI design and algorithm.
- Validation exercise #03: involves a Human-in-the-loop Real Time Simulation where ASTAIR's end-users and stakeholders evaluated ASTAIR's solution. It also involved the Fast Time Simulation for more technical objectives evaluation. Finally, a workshop was organised to conclude on the concept (use cases) and performance areas (human performance, safety, etc.). These activities provided the results, which are linked against the ASTAIR's validation objectives.

The ASTAIR solution, completed the three validation exercises, has reached TRL1, demonstrating initial technical and operational feasibility and potential benefits for ground operations. Simulations showed more predictable and efficient operations. Human-AI collaboration was effective, with high trust reported, despite the recommended improvements for the HMI design and manual editing functions. While the system enables safe fallback to manual human control, clearer task delegation and AI transparency, especially on rationales behind decision-making process, and suggestions, remain essential to address both operational and liability areas. The results confirm that the basic principles of the concept are feasible and relevant for its intended operational environment. Several limitations remain to be addressed in future work, including integration of algorithmic components with the operational HMI, improved AI decision transparency, expanded stakeholder involvement, and clarification of procedural and liability aspects as the concept progresses towards higher maturity levels.

2 Introduction

2.1 Purpose of the document

This document provides the Exploratory Research Report - ERR for SESAR Solution ASTAIR at a TRL0 towards TRL1 level of maturity. It describes the results of the validation exercise TVAL.03.0-ASTAIR-TRL1: Final assessment, defined in the ASTAIR Exploratory Research Plan (ERP). Note that the results from TVAL.01.0-ASTAIR-TRL1 and TVAL.02.0-ASTAIR-TRL1 are reported in the ***D1.2 - Workshops report*** deliverable[2].

2.2 Intended readership

This document is intended for various stakeholders in the Air Traffic Management (ATM) community at large, especially those involved in the SESAR Programme. These include:

- ASTAIR consortium members who prepare and execute the validation activities.
- SESAR 3 JU programme management, and related SESAR 3 projects (CODA, TRUSTY, etc.).
- Academic research and Industry research who wish to learn about the validation activities behind the ASTRA solution.

2.3 Background

The project ASTAIR is not a direct continuation of its previous stage, as the starting maturity level is TRL0. Nevertheless, it does consider the concepts, tools and results from the following projects[3]:

Project Title	Project Description
SESAR EXPLORATORY RESEARCH PROJECT AEON - Advanced Engine Off Navigation (*completed*)	AEON defined a new concept of operations to make best use of green taxiing techniques; specifically, TaxiBots, WheelTugs, e-Taxi, and single-engine taxiing were investigated to address airport ground operations at long to medium-planning and execution phases. ASTAIR is building on AEON Path planning algorithms.
SESAR CODA - Controller adaptative Digital Assistant	The CODA project aims at developing a system in which hybrid human-machine teams collaboratively perform tasks. ASTAIR and CODA do not share the same approach on Human Automation Teaming, especially in the use on neurophysiological measures, nevertheless some questions on delegation strategies may be addressed similarly.

[2] This approach was agreed with the SJU.

[3] More details on the relevant projects and concepts are defined in ASTAIR D3.1 Initial Concept Outline, Chapter 3.2.2.1 Integration with other solutions.

EVOLVE: Motion planning and control in the safety-critical situations (NWO Open Technology Programme, project 18484 (*completed*)	EVOLVE proposes to use an enhanced physics and data-based learning approach to the control of automated driving hazardous driving scenarios known as "edge cases" where representative data are statistically rare. The developed control algorithms will handle and guarantee safety during evasive manoeuvres for collision avoidance, something that current automated driving cannot guarantee. Motion planning and control models that were implemented in AEON and eventually in ASTAIR.
OWHEEL: Benchmarking of Wheel Corner Concepts Towards Optimal Comfort by Automated Driving (EU H2020-MSCA-RISE-2019 OWHEEL, project 872907)	The project OWHEEL aims at the development and evaluation of new concepts of automotive wheel corners as crucial elements of future vehicle architecture tailored to provide an optimal comfort during automated driving. The main goal of the OWHEEL project is to perform a deep analysis and provide on its basis the recommendations for future automated vehicle architecture.
SESAR 2020 exploratory research project Take Control (TaCO)	TaCo aims to define an automated system sufficiently powerful to both accomplish complex tasks involved in the management of surface movements in a complex airport and self-assess its own ability to deal with non-nominal conditions. TaCo allows Air Traffic Controllers to progressively create and tune automation with visual constructs that also assist them in understanding the behaviours, hence facilitating the handover if required. The concept of human automation teaming in ASTAIR project is developed upon the results of Take Control (TaCO)
SESAR TAM - Total Airport Management (PJ04 TAM, grant 733121)	TAM is interesting for ASTAIR development at several levels. First because centralization and automation of ground movement promoted in ASTAIR follows the same philosophy as PJ04 TAM, but also because PJ04 investigated the usage of AI for routing. The developed models proposed to the operational solutions and decision-makers decided whether the proposed solution will be applied. This corresponds to the level 1B (in reference to EASA level of automation). ASTAIR will go further into looking for conflict-free routing.
SESAR TRUSTY – TRUStworthy inTellingent sYstem for remote digital tower	The overall goal of TRUSTY is to provide adaptation in the level of transparency and explanation to enhance the trustworthiness of AI-powered decisions in the context of Remote digital towers (RDT). TRUSTY and ASTAIR will most probably share some problematics concerning human centric AI and human AI teaming, thus staying closely in touch will be fruitful for the project.

Table 1: Relevant results from previous projects that will be fed into the ASTAIR project

2.4 Structure of the document

This Exploratory Research Report (ERR) is structured to provide a comprehensive overview of the final validation activities conducted within the ASTAIR project and the results obtained.

Co-funded by the European Union

Section 1 – Executive summary offers a concise overview of the report, presenting the key objectives, activities, and findings from Validation Exercise #03.

Section 2 – Introduction outlines the purpose, target readership, background context, and structure of the document, along with the glossary of terms and acronyms.

Section 3 – Context of the exploratory research report places the ERR in the wider ASTAIR project framework, summarising relevant elements from the Exploratory Research Plan (ERP) and highlighting any deviations.

Section 4 – Validation results present the outcomes of the final validation activity, with consolidated findings structured by validation objective, along with a confidence assessment.

Section 5 – Conclusions and recommendations draw together the main findings, assesses the progress towards TRL1, and puts forward recommendations for further development.

Section 6 – References provides both reference and applicable documentation for ASTAIR project.

Annexes provide supplementary material:

- Appendix A – Validation Exercise #03 report
- Appendix B - Liability Assessment
- Appendix C - RTS Solution Scenario 2 Description

Note that the results from EXE01 and EXE02 (seen in the table below) **are reported in the *D1.2 - Workshops report* deliverable.**

The following table summarises the validation exercises that will be performed during the ASTAIR project.

ID	High-level validation objective	Rationale	Validation method	Related exercise
1	Validate concept, scope and use cases	Form the basis for the design and development tasks	Workshop with stakeholders	TVAL.01.0-ASTAIR-TRL1
2	Validate solution design and implementation	Develop the final solution design and implementation procedures of ASTAIR. Refine the Use Cases.	Workshops with users	TVAL.02.0-ASTAIR-TRL1
3	Validate ASTAIR solution	Compare results against the validation objectives	HITL RTS, FTS, Final Workshop	TVAL.03.0-ASTAIR-TRL1

Table 2: ASTAIR's three Validation Exercises

2.5 Glossary of terms

Term	Definition	Source of the definition
Airport Moving Map (AMM)	It is a digital display system for pilots that shows the aircraft's current position in the airport surface map	Jeppesen

Human-AI collaboration screen	The display utilized in the Human-Machine Interface (HMI) featured three distinct levels of collaboration: AI (the aircraft is autonomously managed by the system), Collaboration (the aircraft is jointly managed by the human operator and the AI system), Human (human operator assumes full control of the aircraft, enabling manual adjustments such as route changes). Each mode is designed to facilitate varying degrees of interaction between the human operator and the automated system. This screen was integrated within the Supervision interface.	SESAR ASTAIR
Inspection interface	It is the right screen of the HMI, in which the participants could see future trajectories computed by the AI agent	SESAR ASTAIR
Solution Scenario 1	It is the HMI developed by ENAC. It has the Supervision and Inspection interfaces, and it is an Adaptive Automation Level approach.	SESAR ASTAIR
Solution Scenario 2	It is the HMI developed by TUD. It uses an Automation Level 3 approach and MAS (Multi Agent System) algorithms.	SESAR ASTAIR

Table 3: glossary of terms

2.6 List of acronyms

Acronym	Description
A-CDM	Airport Collaborative Decision Making
A/C	AirCraft
A-SMGCS	Advanced Surface Movement Guidance & Control System
AI	Artificial Intelligence
ALDT	Actual LanDing Time
AMM	Airport Moving Map
ASTAIR	Auto-Steer Taxi at AIRport
ATCO	Air Traffic Controller
ATM	Air Traffic Management
CAP	Capacity
CBA	Cost-Benefit Analysis
CRT	CRiTerion
DES	Digital European Sky
E-OCVM	European Operational Concept Validation Methodology

ECAC	European Civil Aviation Conference
EHAM	ICAO code for Amsterdam Schiphol Airport
ENV	Environment
ERR	Exploratory Research Report
EXE	Exercise
FRD	Functional Requirements Document
GA	Grant Agreement
GDPR	General Data Protection Regulation
GND ATCO	Ground Controller
HCD	Human-Centred Design
HCI	Human-Computer Interaction
HE	Horizon Europe
HITL	Human In The Loop
HMI	Human-Machine Interface
HP	Human Performance
ID	Identifier
KPA	Key Performance Area
KPI	Key Performance Indicator
MAS	Multi Agent System
MET	Move Engine Taxi
OBJ	OBJective
OSED	Operational Service and Environment Description
PI	Performance Indicator
R&I	Research & Innovation
RDT	Remote Digital Tower
RMO	Runway Mode of Operations
RTS	Real-Time Simulation
SAF	Safety
SART	Situation Awareness Rating Technique
SESAR	Single European Sky ATM Research
SESAR 3 JU	SESAR 3 Joint Undertaking
SUS	System Usability Scale

EUROPEAN PARTNERSHIP
Co-funded by the European Union

TA	Transversal Area
TET	Towed Engine Taxi
TFM	Tug Fleet Manager
TRL	Technology Readiness Level
TVAL	Test Validation
TWR ATCO	Tower Controller
UC	Use Case

Table 4: list of acronyms

3 Context of the exploratory research report

3.1 Project / SESAR solution ASTAIR: a summary

The SESAR ASTAIR Solution is a support tool to fully automate the supervision of airport ground operations. ASTAIR introduces automated taxiing to improve operational efficiency and reduce delays. It leverages the A-CDM (Airport Collaborative Decision Making) and A-SMGCS (Advanced Surface Movement Guidance and Control System) frameworks to create more predictable ground operations and alleviate operator workloads. The ASTAIR tool consists of an interface that makes use of algorithms to autonomously manage vehicle movements on the airport surface, providing the controller with enough flexibility to locally tweak the algorithm rules to cope with operational events.

ASTAIR aims to promote a cohesive operational environment that integrates manual and autonomous functionalities, enhancing mainly operational efficiency and sustainability in managing engine-off and conventional taxiing operations across major European airports, thus augmenting the capacity of airport ground operations while reducing the impacts on human workload and the environment.

The use of a human-centred approach promotes a coordinated collaboration between human-controlled and automated processes, drawing on operators' expertise to control and engage with the automation (AI) at varied levels, thereby ensuring the optimisation of the collaboration between humans and AI within the complexities of taxi management and control operational tasks.

The exploratory research activities reported in this document focused on assessing the following elements of the ASTAIR solution:

- The feasibility and usability of the HMI in enabling effective human–AI collaboration.
- The capability of the algorithm to generate reliable, safe, and efficient path and motion planning suggestions.
- The solution's operational acceptability in realistic engine-off taxiing scenarios.
- Impacts on human performance, including workload, trust, and situational awareness.
- Potential implications for safety, capacity, environmental sustainability, and liability.

These aspects were explored through Validation Exercise #03, comprising a Human-in-the-Loop Real-Time Simulation (HITL RTS), supported by Fast-Time Simulations (FTS), and a Final Workshop with stakeholders. The objective was to collect evidence on the solution's maturity and validate its core principles in line with the expectations for TRL1.

3.2 Summary of the exploratory research plan

3.2.1 Exploratory research plan purpose

The Exploratory Research Plan (ERP) defined the validation approach for the ASTAIR concept, with the aim of progressing its maturity from TRL0 to TRL1 (Basic principles observed). The validation strategy outlined in the ERP focused on assessing the feasibility and added value of introducing automation to airport ground operations, particularly in support of engine-off taxiing procedures.

The validation included three exercises, of which the final one, Validation Exercise #03, is the focus of this report. The operational environment selected for the development of use cases and validation activities covered major European hub airports, with particular emphasis on apron and ground movement operations. Paris-Charles de Gaulle (CDG), Amsterdam Schiphol (AMS), and Frankfurt (Fraport) airports were chosen as reference environments due to their complexity and strategic approaches to sustainable taxiing, ranging from hybrid engine-off operations to full towing concepts. These airports offered relevant and diverse operational contexts in which to explore automation use cases.

The technical environment defined in the ERP consisted of an algorithmic engine to generate taxi route and motion plans, and a Human-Machine Interface (HMI) designed to support human–automation collaboration. The solution was assessed in a simulated operational context, replicating realistic surface traffic conditions through a Human-in-the-Loop Real-Time Simulation (HITL RTS) supported by Fast-Time Simulations (FTS).

The outcomes of this final validation activity are presented in this Exploratory Research Report (ERR).

3.3 Summary of validation objectives and success criteria

Five validation objectives were addressed as defined in the ASTAIR ERP [3]:

Objective 1

Identifier	OBJ-ASTAIR-ERP-01
Objective	To assess the operational feasibility of the ASTAIR concept.
R&I Need	Adapt intelligent systems to operators' mode of operation
Title	Operational Feasibility
Category	<safety>, <operational feasibility>, <human performance>

Identifier	Success Criterion
CRT-ASTAIR-TRL1-ERP-01.01	Assess that the new ASTAIR procedures and tools are operationally feasible in regards to **pilot's** operating methods based on the feedback.
CRT-ASTAIR-TRL1-ERP-01.02	Assess that the new ASTAIR procedures and tools are operationally feasible in regards to **ATCO**'s operating methods based on the feedback.
CRT-ASTAIR-TRL1-ERP-01.03	Assess that the new ASTAIR procedures and tools are operationally feasible in regards to **Ground Operator**'s operating methods based on the feedback.

Objective 2

Identifier	OBJ-ASTAIR-ERP-02
Objective	Evaluate the collaboration between human-controlled and automated processes/AI.
Title	Human-Machine Collaboration
R&I Need	Collaboration between human-controlled and automated processes/AI

EUROPEAN PARTNERSHIP
Co-funded by the European Union

Category	<performance>, <safety>, <operational feasibility>, <human performance>, <acceptability>, <liability>

Identifier	Success Criterion
CRT-ASTAIR-TRL1-ERP-02.01	Degree of Collaboration - **Teamwork**: Measure the effectiveness of interaction between human operators and the automated system during taxi management tasks based on the operators' feedback.
CRT-ASTAIR-TRL1-ERP-02.02	Integration Flexibility - **Task distribution**: Assess the system's ability to accommodate diverse operator preferences and operational requirements through flexible integration options based on the operators' feedback.
CRT-ASTAIR-TRL1-ERP-02.03	Assess that the **logical consistency** across manual and automated control is ensured based on the operators' feedback.
CRT-ASTAIR-TRL1-ERP-02.04	**Assess the Liability** impact of innovations.*

*Liability CRT supported by the following sub-criteria/metrics:

CRT-ASTAIR-TRL1-ERP-02.**04.01**	**Liability** impact of innovations: Identification of key new liability risks for all actors and stakeholders involved in defining, developing, and implementing the concept, according to the level of definition achieved at various validation stages.
CRT-ASTAIR-TRL1-ERP-02.**04.02**	**Liability** impact of innovations: Identification of suitable measures in design, organisation, and policy to mitigate identified risks.
CRT-ASTAIR-TRL1-ERP-02.**04.03**	**Liability** impact of innovations: Positive feedback from AB stakeholders on the proposed concept or suggestions for alternative enhancements.
CRT-ASTAIR-TRL1-ERP-02.**04.04**	**Liability** impact of innovations: Ensuring that the concept does not introduce unacceptable liability risks for actors and stakeholders.

Objective 3

Identifier	OBJ-ASTAIR-ERP-03
Objective	Assess the operators' controlling and engaging with the automation at diverse levels.
Title	Interaction with different automation levels
R&I Need	*Operator's controlling and engaging with the automation at diverse levels.*
Category	<performance>, <human performance>

Identifier	Success Criterion
CRT-ASTAIR-TRL1-ERP-03.01	**Level of Operator Engagement**: Assess the extent to which operators actively interact with the automated system and utilise its features to enhance operational efficiency.[4]

CRT-ASTAIR-TRL1-ERP-03.02	Customisation options - **Ability to effectively control and engage with the automation** before and during the operation: Measure the range and effectiveness of customisation features available to operators for adjusting system behaviour and settings.[5]

Objective 4

Identifier	OBJ-ASTAIR-ERP-04
Objective	Assess the HMI / interactive tools and adaptive AI algorithms supporting the operators.
Title	Usable HMI and Interactive Tools
R&I Need	*HMI / interactive tools, adaptive AI algorithms*
Category	<performance>, <safety>, <operational feasibility>, <human performance>, <acceptability>

Identifier	Success Criterion
CRT-ASTAIR-TRL1-ERP-04.01	**HMI Usability:** Assess the usability of the HMI and interactions with the interactive tools, based on operators' ability to quickly understand and navigate the interface.
CRT-ASTAIR-TRL1-ERP-04.02	**Decision-making support**: Measure the effectiveness of interactive tools in providing operators with relevant information and assistance for making real-time decisions during ground operations.
CRT-ASTAIR-TRL1-ERP-04.03	**Tools performance:** Evaluate the tools performance and the impact on the efficiency of operators' interactions with the HMI and interactive tools.

Regarding the fifth validation objective above, there was one change in respect to Success Criteria. The success criterion of OBJ05 - CRT-ASTAIR-TRL1-ERP-05.03, was not addressed in the validations due to unforeseen resource constraints. For further details refer to Chapter 3.4.2.

CRT-ASTAIR-TRL1-ERP-05.03: Assess that the manoeuvrability of tugs and tug-aircraft combinations is improved based on the Fast-Time Simulation results. Objective 05 detailed below, for reference, with CRT-ASTAIR-TRL1-ERP-05.03, that is no longer addressed in this solution.

Objective 5

Identifier	OBJ-ASTAIR-ERP-05
Objective	Assess the Optimized Path & Motion Planning for Efficient Ground Operations.
Title	Optimised Path & Motion Planning
R&I Need	Path Planning
Category	<performance>, <safety>, <environmental sustainability>, <capacity>

Identifier	Success Criterion
CRT-ASTAIR-TRL1-ERP-05.01	Assess that the **airport capacity** is maintained or increased with the new ASTAIR concept based on the operators' feedback and the Fast-Time Simulation results.
CRT-ASTAIR-TRL1-ERP-05.02	**Conflict-free routing:** Evaluate the safety implications of optimised taxi routes, including collision avoidance measures and adherence to operational regulations and guidelines. Conflict-free routing (Conflict detection & resolution) to ensure safety levels remained based on the operators' feedback and simulations data analysis.
CRT-ASTAIR-TRL1-ERP-05.03	Assess that the **manoeuvrability of tugs and tug-aircraft combinations** is improved based on the Fast-Time Simulation results**.**
CRT-ASTAIR-TRL1-ERP-05.04	Assess that the **tugs resource management** is improved through the capacity utilisation of the tugs based on the Fast-Time Simulation results

3.3.1 Validation assumptions

Assumption ID	Assumption title	Assumption description	Justification	Impact Assessment
VA-ASTAIR-TRL1-VALP-01	A-CDM	The ASTAIR's solution will be a standalone tool and will assume that Airport collaborative decision making (A-CDM) is more likely to be benefiting from more automation on ground with engine off taxiing techniques. Airports already have dedicated data sharing infrastructure.	This approach is likely to provide the best balance of scalability, cost-effectiveness, performance, manageability, operational advantages.	High
VA-ASTAIR-TRL1-VALP-02	A-SMGCS	The ASTAIR solution will be a standalone tool and will assume that Advanced Surface Movement Guidance & Control System (A-SMGCS) is more likely to be benefiting from more automation on ground with engine off taxiing techniques.	This approach is likely to provide the best balance of scalability, cost-effectiveness, performance, manageability, operational advantages.	High

Table 5: validation assumptions overview

3.3.2 Validation exercises list

The following traceability table specifies the common elements of the ASTAIR for all three validation exercises. This validation report ERR focuses on the results from the third validation exercise - TVAL.03.0-ASTAIR-TRL1.

[ASTAIR EXEs Trace]

Linked Element Type	Identifier

EUROPEAN PARTNERSHIP
Co-funded by the European Union

<SESAR Solution>	0501
<Project>	ASTAIR
<Sub-Operating Environment>	Medium to large airports
<Validation Objective>	To be filled in per exercise.

Table 6: ASTAIR's common elements for the three validation exercises

[EXE]

Identifier	TVAL.01.0-ASTAIR-TRL1
Title	Initial assessment and review of expectations
Description	This exercise involves a series of workshops to refine the concept, the scope and the use cases of ASTAIR together with its stakeholders: Paris CDG Workshop: 18-19/12/2023 Fraport Airport Workshop: 19/04/2024 Expert Group Workshop: 24/05/2024
KPA/TA addressed	All
Addressed expected performance contribution(s)	N/A
Maturity level	TRL1
Use cases	Three ASTAIR Use Cases, addressing the following topics: Arrival without parking, High level taxi strategy tuning, Automation Failure.
Validation technique	Expert Group (Judgement Analysis)
Validation platform	N/A
Validation location	Online
Start date	01/09/2023
End date	24/05/2024
Validation coordinator	ENAC
Status	<completed>
Dependencies	N/A

Table 7: Validation Exercise #01 description

[EXE #01 Trace]

Linked Element Type	Identifier
<Validation Objective>	OBJ-ASTAIR-ERP-01 OBJ-ASTAIR-ERP-03

Table 8: Validation exercise #01 layout

[EXE]

Identifier	TVAL.**02.0**-ASTAIR-TRL1
Title	Intermediate assessment
Description	Workshops to work on solution design and implementation based on prototypes demo. This exercise is for purposes of enabling users to experience how the ASTAIR's system may work, providing prototypes demo of the solution. End-users provide feedback on the solution design, the quality of the information provided and any pain points or missing functions/aspects.
KPA/TA addressed	All
Addressed expected performance contribution(s)	N/A
Maturity level	TRL1
Use cases	Main ASTAIR Use cases: 1,2,3
Validation technique	Expert Group (Judgement Analysis)
Validation platform	N/A
Validation location	ENAC Toulouse
Start date	01/06/2024
End date	30/06/2024
Validation coordinator	ENAC
Status	<completed>
Dependencies	N/A

Table 9: Validation Exercise #02 description

[EXE #02 Trace]

Linked Element Type	Identifier
<Validation Objective>	OBJ-ASTAIR-ERP-01 OBJ-ASTAIR-ERP-03 OBJ-ASTAIR-ERP-04

Table 10: Validation exercise #02 layout

[EXE]

Identifier	TVAL.**03.0**-ASTAIR-TRL1
Title	Final assessment
Description	This exercise represents the execution of the Human-in-the-loop Real time simulation (HITL), Fast-Time simulation (FTS) and the Final Workshop on the defined use cases scenarios. During the test campaign, all the data will be gathered using a mixed approach using quantitative and qualitative

	methodologies (questionnaires, observations, structured-interviews, debriefing, etc.), also including Human Performance evaluation tools addressing HMI usability, user workload and situational awareness. The data gathered will be analysed using standardised research practice to ensure data reliability. Two types of experimental runs have been conducted - one with the system and one without it (baseline/reference), to evaluate the benefits and impacts of implementing the new tool by comparing system performance and user experience across both simulation scenarios.
KPA/TA addressed	All
Addressed expected performance contribution(s)	N/A
Maturity level	TRL1
Use cases	All ASTAIR Use Cases
Validation technique	Human-in-the-loop Real Time Simulation Fast-Time Simulation (FTS) Final Workshop
Validation platform	ASTAIR platform
Validation location	ENAC simulation facilities
Start date	01/05/2025
End date	31/05/2025
Validation coordinator	DEEP BLUE SRL (ENAC, TUD)
Status	<Completed>
Dependencies	N/A

Table 11: Validation Exercise #03 description

[EXE #03 Trace]

Linked Element Type	Identifier
<Validation Objective>	OBJ-ASTAIR-ERP-01 OBJ-ASTAIR-ERP-02 OBJ-ASTAIR-ERP-03 OBJ-ASTAIR-ERP-04 OBJ-ASTAIR-ERP-05

Table 12: Validation exercise #03 layout

3.4 Deviations

3.4.1 Deviations with respect to the S3JU project handbook

This Exploratory Research Report is compliant with the approach to validate/demonstrate SESAR Solutions, as described in the SESAR 3 JU project handbook.

3.4.2 Deviations with respect to the exploratory research plan (ERP)

1. Omission of CRT-ASTAIR-TRL1-ERP-05.03

One notable deviation from the originally defined ERP concerns validation success criterion CRT-ASTAIR-TRL1-ERP-05.03.

CRT-ASTAIR-TRL1-ERP-05.03: Assess that the manoeuvrability of tugs and tug-aircraft combinations is improved based on the Fast-Time Simulation results.

This specific criterion was not addressed in the validations due to unforeseen resource constraints. In particular, the expert partner responsible for the implementation and analysis of this aspect experienced an extended period of unavailability due to long-term medical leave. Given the limited timeframe for executing the simulations and the specialised nature of the task, it was not feasible to reassign or replicate this activity within the scope of the project without compromising the quality of other planned validation tasks.

This deviation has been acknowledged by the consortium and discussed with the SESAR 3 Joint Undertaking. It was agreed that the justification would be transparently reported in the present report. The exclusion of this objective does not significantly impact the overall assessment of the ASTAIR solution at TRL1, as the remaining validation objectives provide sufficient evidence to support the evaluation of the concept's feasibility, performance, and stakeholder acceptance.

2. Algorithms not demonstrated in integration with the main HMI interface

Another deviation concerned the planned integration of the core algorithms with the ASTAIR Human-Machine Interface (HMI).

While the algorithms were successfully validated through Validation Scenario 2 (see below), they were not demonstrated in integration with the main HMI during the Validation Scenario 1 (see below) in the RTS because of their computational complexity. This was due to unforeseen delays in the technical development timeline and resource limitations. As a result, evaluations were performed using standalone modules or mock-up environments in the two separate validation scenarios rather than a fully unified system. This limitation was considered in the interpretation of validation results, particularly concerning HMI usability and decision-making support. Due to this, the validation and the reporting on the RTS week is divided into two validation scenarios described in Chapter Summary of validation exercise #03 validation scenarios in A.1.3.

Therefore, the post-run questionnaire and end of day questionnaire were provided only after the Validation Scenario 1, which enabled better usability and interaction with the system/platform.

Despite these deviations, the consortium considers the overall assessment of the ASTAIR concept at TRL1 to remain valid, with strong evidence gathered across the remaining objectives and validation activities.

EUROPEAN PARTNERSHIP
Co-funded by the European Union

4 Validation results

4.1 Summary of project / SESAR Solution ASTAIR validation results

Table below shows the summary of results compared to the success criteria identified within the VALP per exercise validation objective as well as the status according to the following criteria:

- OK: the validation objective achieves the expectations (the exercise results achieve the success criteria);
- NOK: the validation objective does not achieve the expectations (the exercise results do not achieve the success criteria);
- POK (Partially OK): the validation objective achieves the expectations to a certain extent. The reasons why the validation objective is not fully achieved shall be clearly recorded in the table below.

The validation confirmed the initial feasibility of the ASTAIR concept in supporting the automation of airport ground operations. **While some success criteria were partially OK (POK), this is mainly due to the current maturity level (TRL1)**, the absence of pilot-in-the-loop validation and the integrated algorithmic modules into the main HMI interface not validated in the RTS. These aspects are consistent with the exploratory nature of the research and do not significantly affect the overall outcome. The system demonstrated effective human-AI collaboration, enhanced predictability, improved operational efficiency and conflict-free routing. The minor limitations identified do not compromise the positive evaluation of the concept. Therefore, the benefits demonstrated across the validation activities prevail and all **validation objectives are considered achieved - OK.**

SESAR solution validation objective ID	SESAR solution validation objective title	SESAR solution success criterion ID	SESAR solution success criterion	SESAR solution validation results	SESAR solution validation success criterion status	SESAR solution validation objective status
OBJ-ASTAIR-ERP-01	To assess the operational feasibility of the ASTAIR concept.	CRT-ASTAIR-TRL1-ERP-01.01	Assess that the new ASTAIR procedures and tools are operationally feasible in regard to pilot's operating methods based on the feedback.	The ASTAIR concept shows promise for pilot use, especially with datalink communications, but full feasibility remains unconfirmed due to limited pilot involvement. Further testing with pilots is needed.	POK	OK
		CRT-ASTAIR-TRL1-ERP-01.02	Assess that the new ASTAIR procedures and tools are operationally feasible in regard to ATCO's operating methods based on the feedback.	ATCOs found the automation tools useful but faced challenges with interface design, cognitive load, and unexpected events. Further research is needed to improve tool integration and support during automation failures.	OK	
		CRT-ASTAIR-TRL1-ERP-01.03	Assess that the new ASTAIR procedures and tools are operationally feasible in regard to Ground Operator's operating methods based on the feedback.	Lower levels of automation were seen as feasible and helpful, while level 3 automation needs further evaluation, especially for vehicle drivers. ASTAIR shows potential for integration with existing tools like the Tug Fleet Manager.	OK	
OBJ-ASTAIR-ERP-02	Evaluate the collaboration between human-controlled and automated processes/AI	CRT-ASTAIR-TRL1-ERP-02.01	Degree of Collaboration - Teamwork: Measure the effectiveness of interaction between human operators and the automated system during taxi management tasks based on the operators' feedback.	Participants found the ASTAIR system supported effective human-AI collaboration, with clear task understanding and smooth integration into workflows, though challenges remained in grasping the AI's prioritization of alerts	OK	

SESAR solution validation objective ID	SESAR solution validation objective title	SESAR solution success criterion ID	SESAR solution success criterion	SESAR solution validation results	SESAR solution validation success criterion status	SESAR solution validation objective status
		CRT-ASTAIR-TRL1-ERP-02.02	Integration Flexibility - Task distribution: Assess the system's ability to accommodate diverse operator preferences and operational requirements through flexible integration options based on the operators' feedback.	Most participants found ASTAIR adaptable, supporting shared situational awareness, but the division of responsibilities between human operators and the AI was sometimes unclear.	OK	OK
		CRT-ASTAIR-TRL1-ERP-02.03	Assess that the logical consistency across manual and automated control is ensured based on the operators' feedback.	Operators grew to trust and collaborate with the AI but faced some inconsistencies and unclear roles. Enhancing alignment, contextual info, and pilot communication is needed for better human-AI teamwork	POK	
		CRT-ASTAIR-TRL1-ERP-02.04	Assess the Liability impact of innovations.	The ASTAIR concept carries moderate liability risks mainly around AI reliability and task delegation between humans and AI. Key recommendations include improving AI transparency, clarifying roles, and strengthening procedures and training to ensure safe and accountable operations.	OK	

EUROPEAN PARTNERSHIP
Co-funded by the European Union

SESAR solution validation objective ID	SESAR solution validation objective title	SESAR solution success criterion ID	SESAR solution success criterion	SESAR solution validation results	SESAR solution validation success criterion status	SESAR solution validation objective status
OBJ-ASTAIR-ERP-03	Assess the operators' controlling and engaging with the automation at diverse levels.	CRT-ASTAIR-TRL1-ERP-03.01	Level of Operator Engagement: Assess the extent to which operators actively interact with the automated system and utilise its features to enhance operational efficiency.	The participants demonstrated a high level of engagement throughout the validation runs, with minor decrease from Run 1 to Run 2 of the RTS. They all expressed positivity towards the system, even if sometimes they felt like having a passive role more than a proactive role, especially during routine tasks.	OK	OK
		CRT-ASTAIR-TRL1-ERP-03.02	Customisation options - Ability to effectively control and engage with the automation before and during the operation: Measure the range and effectiveness of customisation features available to operators for adjusting system behaviour and settings.	Customisation options might have some limitations and constraints, which have been addressed, even though it should be taken into consideration that ASTAIR project targets TRL1.	POK	
OBJ-ASTAIR-ERP-04	Assess the HMI / interactive tools and adaptive AI algorithms supporting	CRT-ASTAIR-TRL1-ERP-04.01	HMI Usability: Assess the usability of the HMI and interactions with the interactive tools, based on operators' ability to quickly understand and navigate the interface.	The HMI was generally found user-friendly and usable, with most participants quickly understanding its functions within very limited training. Some improvement areas have been identified.	POK	

SESAR solution validation objective ID	SESAR solution validation objective title	SESAR solution success criterion ID	SESAR solution success criterion	SESAR solution validation results	SESAR solution validation success criterion status	SESAR solution validation objective status
	the operators.	CRT-ASTAIR-TRL1-ERP-04.02	Decision-making support: Measure the effectiveness of interactive tools in providing operators with relevant information and assistance for making real-time decisions during ground operations.	Participants found the system supportive and timely for decision-making, showing good situational awareness during exercises. While some tools need improvements, overall, the system effectively aided decision-making in air traffic control.	OK	OK
		CRT-ASTAIR-TRL1-ERP-04.03	Tools performance: Evaluate the tools performance and the impact on the efficiency of operators' interactions with the HMI and interactive tools.	Participants trusted the AI's performance in traffic integration and conflict inspection, especially in Solution Scenario 2 with higher level of automation, and generally agreed with its recommendations. However, they expressed a need for more information regarding the AI's decision-making rationale.	POK	
OBJ-ASTAIR-ERP-05	Assess the Optimized Path & Motion Planning for Efficient	CRT-ASTAIR-TRL1-ERP-05.01	Assess that the airport capacity is maintained or increased with the new ASTAIR concept based on the operators' feedback and the Fast-Time Simulation results.	Fast-time simulations at Schiphol showed that the ASTAIR system maintains or improves runway capacity and reduces taxi times compared to historic data. It also enhances slot compliance and optimizes runway usage during peak periods.	OK	

EUROPEAN PARTNERSHIP
Co-funded by the European Union

SESAR solution validation objective ID	SESAR solution validation objective title	SESAR solution success criterion ID	SESAR solution success criterion	SESAR solution validation results	SESAR solution validation success criterion status	SESAR solution validation objective status
	Ground Operations.	CRT-ASTAIR-TRL1-ERP-05.02	Conflict-free routing: Evaluate the safety implications of optimised taxi routes, including collision avoidance measures and adherence to operational regulations and guidelines. Conflict-free routing (Conflict detection & resolution) to ensure safety levels remained based on the operators' feedback and simulations data analysis.	The MAS model, calibrated with historic data, accurately plans aircraft paths by considering detailed airport operations and safety requirements. Fast-time simulations showed the routing algorithm effectively produces conflict-free trajectories	OK	OK
		CRT-ASTAIR-TRL1-ERP-05.04	Assess that the tugs resource management is improved through the capacity utilisation of the tugs based on the Fast-Time Simulation results	Simulations show that increasing the number of Electric Towing Vehicles (ETVs) reduces total fuel consumption of outbound taxiing aircraft by up to 38% for RMO North and 28% for RMO South, with diminishing returns beyond certain fleet sizes	OK	

Table 13: summary of validation exercises result

EUROPEAN PARTNERSHIP

Co-funded by the European Union

4.2 Detailed analysis of project / SESAR solution validation results per validation objective

This section provides an analysis of the results obtained from the validation exercise, broken down per the validation objective. It should be noted that normally this chapter provides a consolidated analysis on the solution level; however, given that this report reports only on one final validation exercise (03), this section provides both the solution-level and the exercise-level analysis.

4.2.1 OBJ-ASTAIR-ERP-01 Results

Objective: **To assess the operational feasibility of the ASTAIR concept.**

Title: Operational Feasibility

R&I Need: Adapt intelligent systems to operators' mode of operations

Result:

Overall, the results were OK.

The operators feedback was addressed both during validation runs and workshops. The ASTAIR concept and procedures were widely accepted and feasible for pilots, ATCOs and ground operators, even though there are some further improvements and recommendations that need to be addressed in future research. They, indeed, both had positive and negative aspects, which have been considered in the recommendations.

The biggest limitation comes from pilots not being involved in the validation runs and, therefore, missing their feedback on the tools.

CRT-ASTAIR-TRL1-ERP-01.01

Pilot – Assess that the new ASTAIR procedures and tools are operationally feasible in regard to **pilot's** operating methods based on the feedback.

Result:

The operational feasibility of the ASTAIR concept from the pilot's perspective was partially addressed during the validation runs. It was, indeed, primarily assessed through scenario design and feedback collected during workshops, especially the Final Workshop of ASTAIR that has been conducted at the end of May. Although pilots did not directly participate in the Real-Time Simulation, key aspects of their interaction with the system, particularly the implications of datalink communications and routing predictability, were discussed with operational experts and pilot representatives.

Stakeholders acknowledged that the concept aligns with current trends in cockpit digitalisation and collaborative surface management. The removal of routine voice communication (radio) in favour of datalink for route and clearance transmission was generally well received, but instructions should be clear, timely, and integrated into cockpit avionics (e.g., airport moving maps). However, concerns were

Co-funded by the European Union

raised about pilot's situational awareness and compliance in the absence of verbal confirmation. During the Final Workshop, stakeholders recommended that any future implementations ensure visual clarity and confirmation of instructions within cockpit systems.

Limitation and recommendation:

While initial feedback from stakeholders suggests that the concept is operationally feasible from the pilot's point of view under certain enablers (e.g., cockpit HMI, datalink, reliable A-CDM), direct pilot-in-the-loop (as an end user) validation remains a limitation of this study. Further empirical validation with pilots will be needed in future concept development to confirm full feasibility.

CRT-ASTAIR-TRL1-ERP-01.02

ATCO – Operational feasibility: Evaluate the concepts that foster collaboration between human operators and autonomation in the ASTAIR based on ATCO feedback.

Result:

During the runs of the final validation, the three ATCOs participants got more and more comfortable using the automation supervision and inspection interfaces of the Solution Scenario 1 They appreciated the general ease of learning of the supervision and Inspection tools use. One of the participants could even already picture themselves with that type of tools in the control tower.

Participants, during the run that validated use case 1 (Departure and Arrival with TaxiBot), also mentioned that they liked the fact that the AI plans were following standard traffic rules (i.e. normal traffic procedures) on the airport, which contributed to the trust they gave to the AI. In fact, the participants were supportive of collaborating with an autonomous agent that could give clearances for direct flight routes with no hazard. These results provide positive feedback for the use case 1 (Normal operations), demonstrating that the ASTAIR tools are expected to provide support on the top of the current operations and procedures in managing traffic.

The automation plan Inspection interface did not come without limitation. Although the participants found the Inspection interface useful during the runs, the location of the Inspection interface on another screen was not optimal. The participants could not inspect automation plans as much as they wished, being more focused on the real time traffic on the A-SMGCS. Inspecting the automation plans for upcoming hazardous situations took some time while monopolising participants' visual attention. In addition, one participant seemed confused about the different time spaces implemented in the tools. Going back and forth from the A-SMGCS real time traffic to the Inspection interface future traffic was conceptually difficult for this participant.

Finally, a participant got overloaded, during the validation of the use case 8 (Departing aircraft with technical issues), when a departing flight taxiing to the runway threshold requested to taxi back to its parking stand because of a sick passenger. The participant had to stop all surrounding taxiing flights to provide a safe route back to the apron, which caused a significant congestion on the platform. This highlights the challenge of handling uncertainty when collaborating with very high level of automation, especially when it comes to disruptions or emergency scenarios, such as the ones described in the before mentioned use cases.

Co-funded by the European Union

Given the TRL of ASTAIR, the tools proposed in this exploratory research are not ready for production yet. Nevertheless, most of the concepts that these tools have substantiated were praised by the participants. Therefore, these concepts should be developed further to offer adequate collaboration between humans and automations for ground control.

On the other hand, Solution Scenario 2 was perceived by Air Traffic Controllers (ATCOs) as less operationally feasible, as the AI system assumed responsibility for many of the routine tasks typically managed by the ATCOs themselves. This causes the ATCOs to feel as a supervisor rather than having an active role. Nevertheless, during the debriefing sessions, participants consistently acknowledged that the AI system managed the traffic in a safe and efficient manner, resulting in an orderly and expeditious traffic flow over the airport surface.

It was also noted, however, that certain elements of the scenario, such as stand allocation, do not usually fall within the ATCOs' scope of responsibilities. The inclusion of such tasks introduced additional workload during the runs.

Many participants reported feeling more engaged and operationally effective in scenarios involving emergencies or conflicts, such as Use case 8 (Departing aircraft with technical issues), Use case 2 (Normal operations with re-scheduling) and Use case 6 (Arriving aircraft with occupied parking). In these situations, they appreciated the AI’s ability to handle routine processes, thereby allowing them to focus their cognitive resources on higher-priority tasks.

A particularly valued feature of the system was the ability to view the AI's future trajectory predictions and planned actions on the same visual layer as the real-time traffic.

Nonetheless, participants also expressed concerns that the scenarios presented an overly idealised representation of AI capabilities, especially for the Use case addressing normal operations (UC1a, UC1b). The simulated environments appeared highly controlled and optimistic, lacking the unpredictability and variability that characterize real-world operations.

Recommendation:

To improve the AI plans Inspection interface, further research is needed to integrate its features into the A-SMGCS to reduce visual spread and potential cognitive load. In addition, research should also address the challenge of working with real time and future time spaces together.

Additional research should address the handover to human operators when automation fails or when unexpected event arises. The automated system should be able to partially support human operators to avoid extreme traffic situations on airport platforms. Research could focus on providing degraded automation solutions that can still support human operators when the automation with the highest level of autonomy fails.

CRT-ASTAIR-TRL1-ERP-01.03

Ground operator – Operational feasibility: indicates perception of the feasibility by the airport operator and vehicle driver (other than tug driver) at the airport.

Result:

Co-funded by the European Union

From the airport operator point of view, the participant and observers reported that the integration of automated processes / AI of the level 1 and 2 for specific tasks is perceived as operationally feasible. For example, the allocation of parking stands or tug fleet allocation as a partially automated process is seen as a decision-making tool to help the airport operator in charge of this task. For vehicle drivers, others then tug one, the impact of change of the reduced use of radio communication between ATCOs and pilots / tug drivers is raised and would need further study. This was addressed during the Final Workshop.

For the Solution Scenario 2, in the RTS, where the radio communication is limited and all instructions are provided via data link, the interface for ground / airport operator is limited. For a tug driver or any other driver on the movement area at the airport, the HMI interface for receiving instructions was not evaluated. Hence at the stage is it too early to conclude on the operational feasibility of level 3 automation developed in ASTAIR in relation to the other vehicle drivers.

Moreover, the Tug Fleet Manager (TFM) solution, designed and developed in AEON (and proposed as a SESAR ATM Solution) could benefit from the ASTAIR solution as well. In the same manner the A-SMGCS radar image can be coupled with ASTAIR supervision HMI, the Tug Fleet Management interface can be used to select an empty tug and analyse its AI computed trajectory, Figure 1 illustrates this.

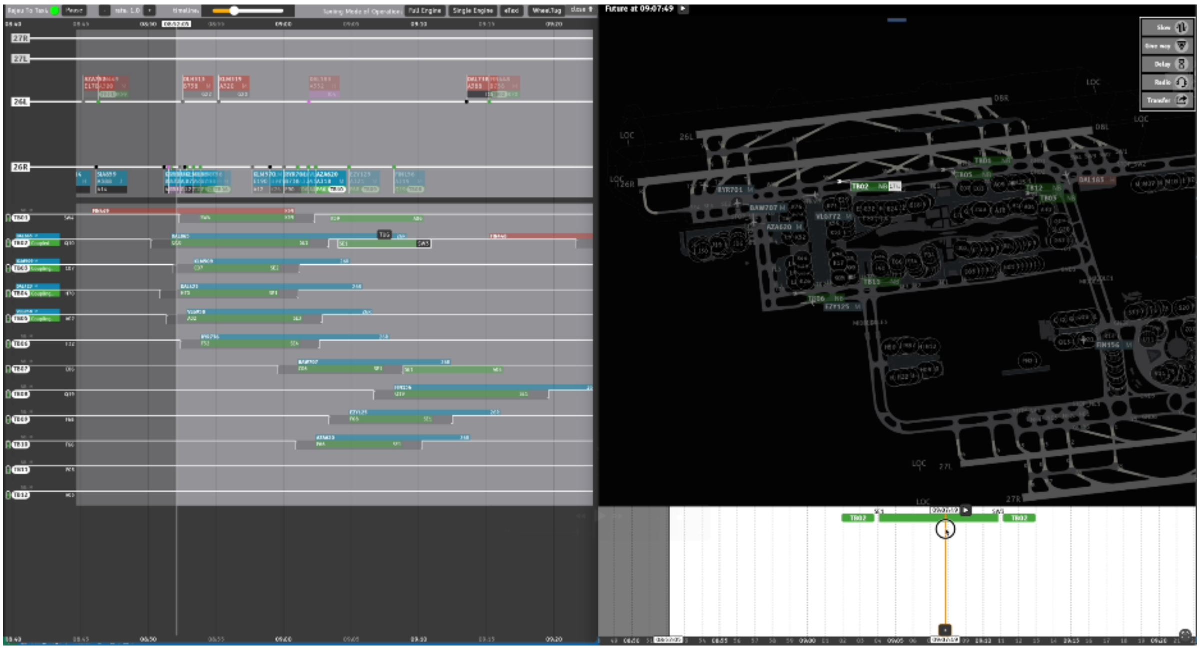

Figure 1: Tug Fleet Manager with inspection HMI

- Left is the TFM HMI.
- The first 4 lines show the incoming traffic sorted by runways (blue are departure and red arrival)
- Under the runways there is one line per TaxiBot showing their mission (green bars show empty tug movement, red/green or blue/green are tug towing an A/C). Not shown here but the AI can also plan for TaxiBot charging period in between missions.
- These missions are initialised with AI plan but can be modified manually by the operator according to operational events.

EUROPEAN PARTNERSHIP
Co-funded by the European Union

- Right is ASTAIR 'future radar image', i.e. the radar image showing the future as it has been computed by AI. The operator can select a tug to investigate its planned movements when empty. The movements when the tug tows an A/C are not shown because it would be entirely under ATCO responsibility then. The planned trajectories for tugs are computed together with all other vehicles by MAS.

This tool was evaluated during the RTS week in a separate session and qualitative feedback was provided by the ground operator, as reported in this criterion.

Furthermore, it was mentioned both by Airport operators and stakeholders at the Final Workshop, that there are some tasks which will need to be done by other actors (e.g. stand allocation is not under ATCO's responsibility, in some specific airports ATCO will have to drive tugs back, etc.). In addition, coordination between Tug Fleet Manager and ATCOs, impact on ATCO workload related to scenarios with different levels of engine-off taxiing techniques (could impact period at which taxiing techniques can/should be used along with environmental consideration), digital communication aspects will need to be investigated, as well as some interfaces (e.g. how an aircraft list to be towed is provided to the tug fleet manager). Digital communication aspects between Tug Fleet Manager and Tug drivers will also need to be addressed in future research and development to increase the maturity of the concept.

Recommendation:

To further evaluate operational feasibility for other stakeholders concerned by the solution: airport operator / vehicle drivers (winter service, firefighting service, wildlife service, airfield inspection drivers etc.). The change management of the solution for different operational services need to be conducted.

4.2.2 OBJ-ASTAIR-ERP-02 results

Objective: **Evaluate the collaboration between human-controlled and automated processes/AI.**

Title: Human-Machine Collaboration

R&I Need: Collaboration between human-controlled and automated processes/AI

Results:

Overall, results were OK.

Participants showed effective collaboration with the ASTAIR system, integrating it into their workflows and maintaining situational awareness. They trusted the AI's assistance but noted some challenges in understanding conflict prioritisation and task distribution, indicating a need for clearer communication from the system on its context and decisions.

Furthermore, participants were optimistic about AI's potential to transform air traffic control by streamlining workflows and reducing routine tasks, though some of them felt their role shifted toward supervision, which may impact overall job satisfaction. This happened especially during Solution Scenario 2, where the AI was taking care of all routine tasks. On the other hand, cognitive workload varied, often improving with system familiarity for both Solution Scenarios.

Co-funded by the European Union

Key recommendations include improving alert prioritisation, clarifying task allocation, enhancing AI decision transparency, and addressing pilot communication to strengthen human-AI collaboration and trust. Liability analysis and recommendations were also addressed (CRT-ASTAIR-TRL1-ERP-02.04).

CRT-ASTAIR-TRL1-ERP-02.01

Degree of Collaboration - Teamwork: Measure the effectiveness of interaction between human operators and the automated system during taxi management tasks based on the operators' feedback.

Results:

The participants' experience highlighted a valid degree of collaboration between human operators and the ASTAIR system. Most of the participants demonstrated the ability to work fluidly with the tool, in both Solution Scenario 1 and 2, integrating its features into their workflow with increasing ease over time. From the perspective of human-AI collaboration, participants reported a clear understanding of task-related information and progress, as well as the level of automation set up during the run of Solution Scenario 1. This finding was consistently supported across the questionnaire responses, debriefing sessions, and direct observations, as most participants predominantly used the Human AI collaboration column of the HMI for most interactions and throughout the different tasks. The system supported dynamic engagement, with operators actively monitoring and, when necessary, intervening in AI processes, indicating an adequate collaboration between human judgment and AI assistance.

At the same time, difficulties in understanding the priorities and time sensitivity of the AI agent's decisions were noted, as evidenced and supported by the questionnaire results (Figure 2).

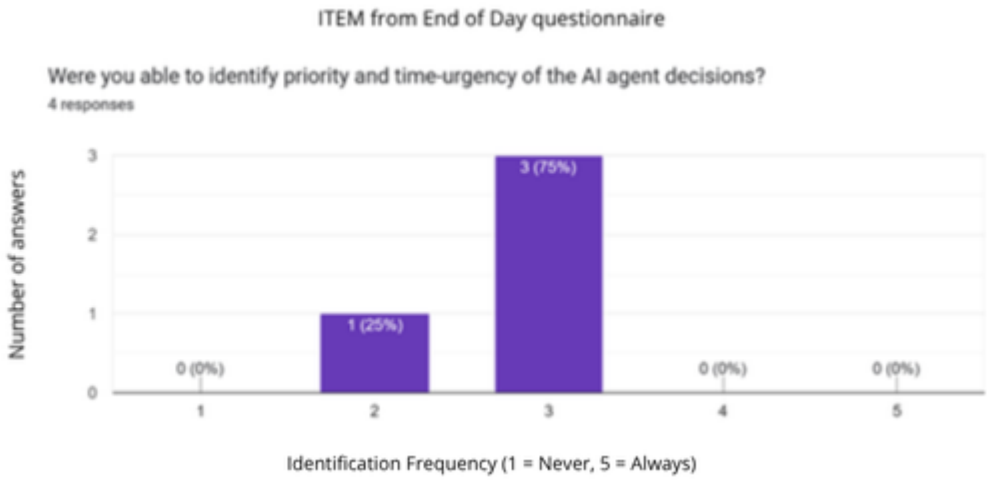

Figure 2: Answer to the question "Were you able to identify priority and time-urgency of the AI agent decisions?" from End of day questionnaire

This difficulty was further confirmed by participants during the debriefing sessions. For example, one participant noted: *"There were some alerts or conflicts that I perceived as less important than others, but the AI agent did not prioritise them based on their importance, which made it challenging for me to understand the time urgency".* This was an important topic also for other ATCOs during Solution Scenario 1 runs.

EUROPEAN PARTNERSHIP
Co-funded by the European Union

Overall, participants expressed a positive impression on the collaborative potential between human operators and the AI agent, noting that it is likely to bring about a significant and constructive transformation in the role of Air Traffic Controllers (ATCOs). This evolution is expected to streamline existing workflows, reduce the burden of routine tasks (this was especially noted during the runs of Solution Scenario 2), and enhance operational efficiency. Participants also acknowledged that such a shift will require targeted adjustments, including the development of updated procedures and specialized training programs, to fully support a smooth and effective transition toward more AI-integrated operations.

Indeed, the results from the Bedford cognitive workload (Figure 3) were in line with this statement. Following the use of the ASTAIR tool during Solution Scenario 1, most participants rated their cognitive workload at level 6, which corresponds to 'little spare capacity', a level of effort that allows limited attention to additional tasks. While this indicates a tolerable reduction in mental spare capacity, individual experiences varied. One participant reported a decrease in cognitive workload between the first and second runs, moving from level 7 to level 6. Notably, the participant attributed this reduction to an increased familiarity with the tool, which made the tasks less cognitively demanding over time. On the other hand, another participant indicated an increase, from level 5 to level 7, therefore entering a category characterized by a non-tolerable reduction in spare capacity. However, the number of aircraft involved was notably high (N=15), and participants were required to conduct TaxiBot operations, tasks that are not part of their usual procedures. This increased operational demand serves as a mitigation factor when assessing cognitive workload. Despite these added challenges, the overall cognitive workload remained acceptable.

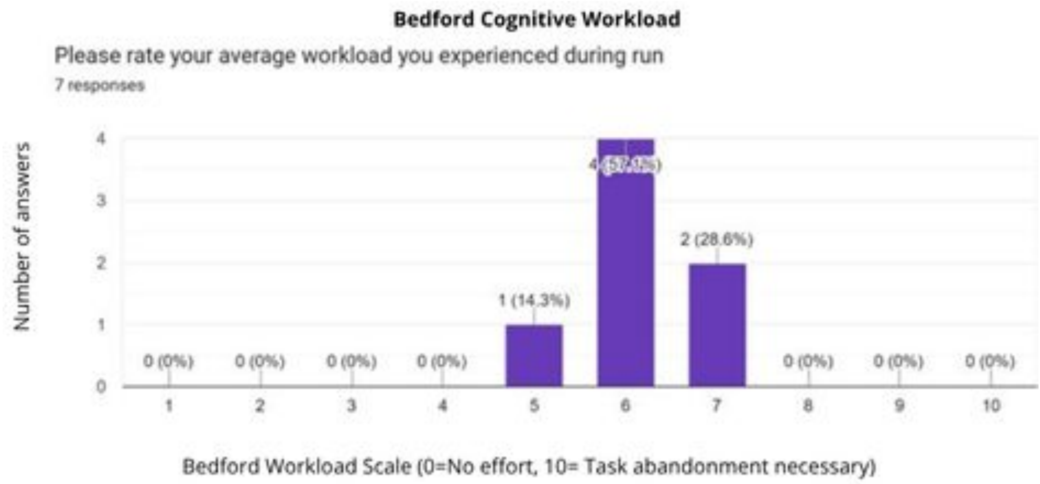

Figure 3: Bedford Cognitive Workload results

Recommendation:

A possible improvement of the system could be to prioritise alerts and conflicts based on the current traffic conditions and the system's traffic predictions. Additionally, the inclusion of labels on strips indicating a conflict may be beneficial for the cognitive workload and situational awareness.

Limitation:

Different participants, especially during the Solution Scenario 2, reported perceiving their role primarily as supervisors, which led to a sense of reduced usefulness compared to current operational practices. This perception may have implications for long-term job satisfaction.

CRT-ASTAIR-TRL1-ERP-02.02

Integration Flexibility - Task distribution: Assess the system's ability to accommodate diverse operator preferences and operational requirements through flexible integration options based on the operators' feedback.

Results:

During the Solution Scenario 1, the information provided supported a shared understanding of the situation for most participants (75%). Indeed, during the debriefing sessions and direct observations, one participant stated that the ASTAIR system demonstrated a high level of integration flexibility, enabling the participant to seamlessly intervene or override simple automated suggestions when desired. This adaptability was particularly evident when the participant intentionally tested the system's response to human inputs, such as rejecting AI-generated routes, not due to their inaccuracy, but to explore the boundaries of manual control. He appreciated that the tool allowed such adjustments without disrupting overall system performance. However, the division of tasks and responsibilities between the human operator and the AI agent was less clearly defined (Figure 4).

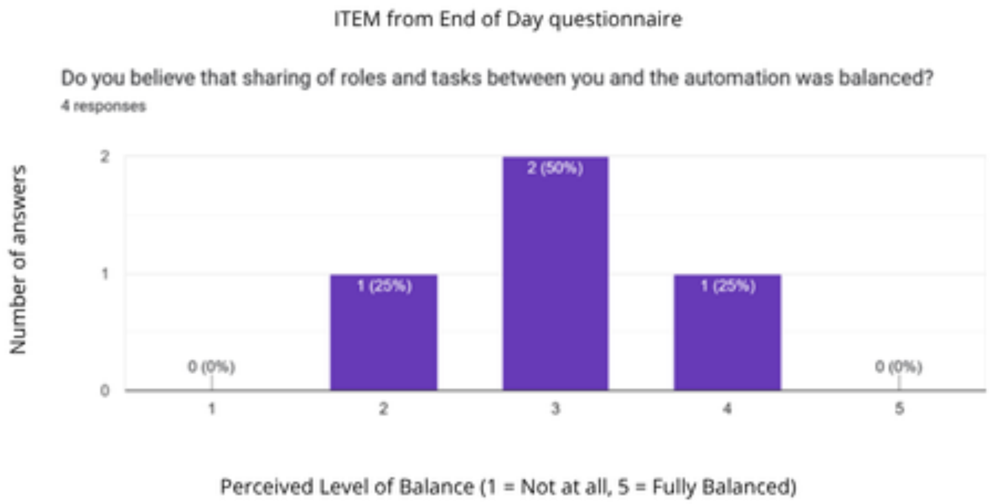

Figure 4: Answer to the question "Do you believe that sharing of roles and tasks between you and the automation was balanced?" from End of day questionnaire

During the debriefings, one participant reported uncertainty regarding which tasks were assigned to the AI agent. While several others perceived the distribution of responsibilities between themselves and the AI agent as not so well-balanced.

Additionally, many participants reported uncertainty regarding the information utilised by the AI to compute their tasks, such as conflict-free route generation Consequently, while they generally trusted that operations were proceeding smoothly, they expressed a preference for receiving more detailed information to enhance their confidence and reassurance. It should be noted, however, that during the Roissy runs (Solution Scenario 1), the AI decisions were based on pre-defined, manually created data rather than actual AI-driven outputs or decisions. As such, these simulated AI decisions may not accurately reflect how a finalised actual AI system would behave. As a possible recommendation, a more thorough definition of these outputs, developed in collaboration with air traffic control (ATC) experts, could help mitigate participant uncertainty in future evaluations or studies.

During Solution Scenario 2, participants observed that the tool demonstrated limited flexibility. One participant suggested that the MAS could offer enhanced support or increased levels of automated coordination, particularly in non-nominal situations such as emergencies.

CRT-ASTAIR-TRL1-ERP-02.03:

Assess that the logical consistency across manual and automated control is ensured based on the operators' feedback.

Results:

Overall, the operators perceived their role as collaborative with the AI agent. Initially, they considered themselves primarily as supervisors, both during Solution Scenario 1 and Solution Scenario 2; however, over time, they became more proactive by actively monitoring and managing conflicts and placing trust in the AI agent to compute conflict-free routes and/or other routine tasks. For most part of the runs during Solution Scenario 1, they reported a shared situational awareness of the traffic alongside the AI agent, as demonstrated below (Figure 5). This obviously helped in decision-making processes over the runs. Indeed, during debriefing sessions they stated that AI's behaviour was consistent with operational rules, particularly during nominal scenarios. This was addressed as a positive remark also during Simulation Scenario 2.

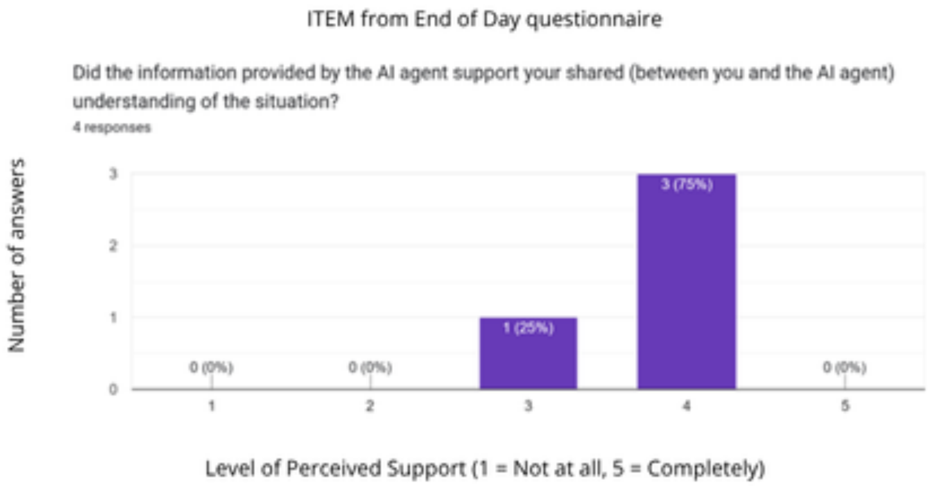

Figure 5: Answer to the question "Did the information provided by the AI agent support your shared understanding of the situation?" from the End of day questionnaire

However, during the debriefings, several participants' remarks underscored the need for further improvements to ensure enhanced logical consistency between human and AI control. While operators generally demonstrated an understanding of the situation and effective collaboration with the AI agent, discrepancies in decision-making processes were observed during Solution Scenario 1 runs. For example, two participants found it necessary to repeatedly verify how the AI agent resolved certain conflicts, and one participant chose to continue using the Human column on the Human-AI collaboration screen to manually adjust or override route decisions. It should be noted, however, that this observation must be considered in light of a key experimental limitation: the AI plans could not be updated in real-time, as a simulated AI was used. This constraint clearly reduced the operators' ability to intervene freely without risking unintended consequences or system disruptions. Therefore, these findings indicate areas where alignment between human and AI control could be improved but also reflect the impact of the testing environment on participants' interactions with the AI.

Nevertheless, when necessary, users demonstrated a good ability, within the system's capacities, to override the decisions and actions of the AI agent (Figure 6). This indicates a high level of consistency between manual and automated control, which is safety critical in situations where the AI malfunctions or makes an error.

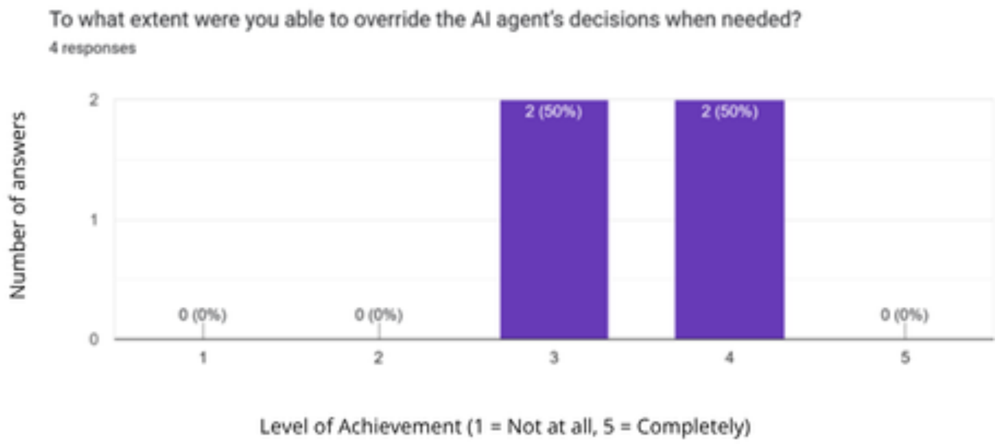

Figure 6: Answer to the question "To what extent were you able to override the AI agent's decisions when needed?" from End of day questionnaire

Participants also noted that the system could be improved by including additional contextual details, such as the priority level or sequencing order of aircraft, to better support decision-making and maintain consistent logic between human judgments and AI outputs.

Furthermore, concerns were raised regarding pilot compliance to the clearance, particularly when communications are conducted via data link rather than voice. Indeed, some participants expressed uncertainty about whether pilots would follow instructions in the absence of verbal confirmation. It was reported by an airline expert present at the validations that it happens in real operations that pilots do not follow strictly speed advisories. In such cases, logical consistency may be compromised if the AI assumes compliance while the human operator remains unsure, underlining the importance of integrating communication status (e.g., frequency acknowledgement) into the decision-support system and other A-SGCS functional block such as Safety Support / Alerting Service (Safety Nets, including the conformance monitoring). Nevertheless, the project assumes (also in the Assumptions chapter that the concept is applied to the environments with A-SMGCS tools in place.

In high-pressure situations, operators prioritise safety to prevent conflicts, which may limit their ability to plan ahead or anticipate future traffic. This focus on immediate safety, combined with the complexity of managing multiple screens and information, can lead to mismatches between human decisions and AI actions, causing inconsistencies in how the system functions. Additionally, one of the air traffic controllers noted that the time frame for potential conflicts was too much in advance (in time) Specifically, conflicts projected to occur 20 minutes in advance were perceived as of limited relevance, since the airports validated (high complexity airports – CDG and Schiphol) experience frequent changes to aircraft flows and high traffic levels and, therefore, their traffic flow can rapidly change.

Regarding the Solution Scenario 2, the participants were all very satisfied with the automated control of the AI, and reported that the AI followed the standard routes. However, one participant mentioned there were some actions (e.g. detour, stopping on taxiway, etc.) that it is up to ATCO's expertise to decide how to manage them.

Recommendation:

The consideration regarding pilots' compliance was addressed during the Final Workshop of the project. In particular, the role of pilots and their perception of information provided by the AI system should be carefully examined, as it may influence their situational awareness and trust in the communication process.

During the Final Workshop the participants also mentioned, as a recommendation for this specific point, that using the Airport Moving Map (AMM) to receive information and clearances could be beneficial for the pilots, because it would simplify the communication, situational awareness and understanding of the pilot.

CRT-ASTAIR-TRL1-ERP-02.04

Assess the Liability impact of innovations.

Liability CRT is supported by the following sub-criteria/metrics:

Results:

The liability assessment has been performed using the Legal Case Methodology (B.2), with the support the Liability Tool, a proprietary asset of Deep Blue[4]. This evaluation builds on the outcomes of the human factor analysis (4.2.34.2.3) and draws from the descriptions of the UCs covered by the validation activities and the respective sequence diagrams[5].

[4] The Liability Tool is an asset proprietary to Deep Blue. The tool consists of a web-based software application that formalises and standardises the steps and activities of the Legal Case methodology, facilitating and supporting the analysis.

[5] ASTAIR (2024). D1.3 - Initial concept outline (https://research.dblue.it/astair/portfolio-items/d1-3-initial-concept-outline/?portfolioCats=2).

Co-funded by the European Union

The results provided high level overview of the possible legal risks associated with the ASTAIR ConOps, in the whole and for each UC, in light of the tasks considered in the proposed sequence diagrams for nominal and non-nominal scenarios. The assessment provides approximate outcomes that focus on the potential liability exposure, in theory, with no reference to likelihood.

The analysis took into account a wide range of scenarios, including the current tasks, the new tasks and the revised tasks. Moreover, the assessment also included the causal dependences among the tasks, intended as the correlation between actions and their consequences. This approach enables a deeper scrutiny of the possible effects of mis-coordination among the actors and systems involved. The focus is on new or revised tasks and their causal dependencies. Current tasks were considered only when affected by novelties introduced by or correlated to the use of the ASTAIR solution.

Against this background, and as summarised in the table below, the ConOps entails a moderate level of exposure to liability risks, with 46% of the scenarios considered potentially involving such risks. As can reasonably be expected, the UCs presenting non-nominal scenarios or a higher number of new tasks are slightly more exposed than others. These results, however, should be read also in light of the level of complexity of the interactions among the actors and the system involved.

Risk / Scenario	UC1.a	UC1.b	UC2	UC3	UC4	UC5	UC8	Tot.
New tasks	15	12	11	13	9	7	5	72
Revised tasks	2	0	1	0	0	7	1	11
Current tasks	2	5	4	7	2	0	4	24
Causal dependencies	15	11	10	4	5	13	10	68
Analysed situations	36	25	32	19	15	28	15	170
Potential liability risks	11	12	13	15	6	15	7	79

Table 14: Overview of possible liability risks related to the ASTAIR ConOps

In light of the above, it should be emphasized that the most frequent legally relevant risk scenarios mainly concern:

- the proper functioning of the AI when it is required to produce reliable inputs for the taxiing operations management process, particularly in terms of AI assurance and operational explainability.
- the delegation of functions from the human operator to the AI system, especially when this transfer of responsibility gives rise to new forms of human-AI interaction involving other operators (e.g., the GND ATCO delegates a task to the AI and informs the PIC they have been transferred to AI).
- and the reassumption of functions by the operator after they have been delegated to the AI. This is especially critical in non-nominal scenarios, such as automation failures or operations involving technical issues.

The rationale underlying these general considerations is detailed in the results related to the sub-objectives presented in the following pages and in Appendix B of this document.

Liability impact of innovations: Identification of key new liability risks for all actors and stakeholders involved in defining, developing, and implementing the concept, according to the level of definition achieved at various validation stages.

Results:

The analysis of liability risks per actor focused on the entities most affected by the introduction of the proposed solutions. These include the AI system (and its provider), the Air Traffic Controllers (especially the GND ACTO), the Pilot in Command (PIC), and the Tug Fleet Manager (TFM). The applicable legal framework for each actor is provided in Appendix B (B.3.1, B.3.2,B.3.3).

Consistent with the approach outlined above, the following table presents the results of the analysis.

Actor	Liability risk	UC1.a	UC1.b	UC2	UC3	UC4	UC5	UC8	Tot.
AI	**Product Liability**	8	8	7	8	7	6	6	**50**
(GND) ATCO	**Professional Liability**	8	4	7	7	3	9	3	**41**
TFM	**Professional Liability**	3	4	0	0	3	7	4	**21**
PIC	**Professional Liability**	4	3	3	2	0	4	0	**16**

Table 15: Overview of liability risks per actor

Product liability

As shown, the AI system (and its provider) results exposed to liability risks in 63% of the total scenarios identified as involving potential liability risks. However, this outcome reflects the specificities of the current legal and regulatory regime of this type of technology. Considering the potential effects of system failure or malfunction, the recurring risks identified generally relate to:

- Possible design defects related to deficiencies in the interface and system architecture, which may result in insufficient support for human oversight (with the risk of compromising the operator's ability to make autonomous and accountable decisions) as well as possible system unexpected behaviours that may contrast with users' reasonable expectations.
- Possible manufacturing defects related to intrinsic vulnerabilities of system, models and datasets due to robustness and cybersecurity issues.
- Possible warning defects related to inadequate communication to deployers of required input data quality, computing resources, system capabilities and limitations, vulnerabilities, or maintenance requirements. In addition, worth to be included also the potential risks emerging in case of insufficient guidance on system usage, interpretation of outputs, detection of system abnormal behaviours, and associated operational protocols.

More details available in Appendix BB.2.4.3.1.

4.2.2.1 Professional liability – ATCO

The liability risk exposure associated with the ATCO (especially the GND ATCO) warrants reporting and brief analysis, as it may be relevant in 52% of the potentially problematic scenarios identified. Based on the current ConOps description and related use case sequence diagrams, GND ATCOs appear particularly susceptible to human errors, including omissions or failures to perform expected duties in most of the scenarios involving their role. These scenarios also include causal dependencies, where the outcome of a preceding task, carried out either by the system or another actor, is brought to the GND ATCO's attention.

This risk profile is primarily attributable to the specific role and responsibilities assigned to GND ATCOs in managing ground operations, including taxiing. Controllers in this position are accountable not only for their own actions but also for those of others, within the scope of the operations they oversee. This finding reflects the broad scope of duties applicable to this role and should be interpreted in conjunction with the results related to the TFM.

More details available in B.3.2.

4.2.2.2 Professional liability – TFM

Although in quantitative terms the TFM results exposed to a lower legal risk (27% of the potentially problematic scenarios considered) the operative interdependence in terms of roles, competencies and tasks with the GND ATCO is crucial. Indeed, the TFM is a relatively new role in ground operations, especially with the tasks and duties envisioned in ASTAIR.

Following the analysis conducted within the AEON project[6], from a legal perspective, if this subject is considered as a dispatcher equivalent figure, the legal risks on the GND ATCOs may increase, due to their accountability duties on the tasks and procedures performed also by other actors under their monitoring. Conversely, if the TFM is intended as an ATCO equivalent figure, the legal risks on the GND ATCO may be more contained, at least qualitatively, and the TFM should be held accountable for the operations under their management.

This latter interpretation is the one supported within the ASTAIR project, with the aim to enable a more effective allocation of responsibilities, particularly in relation to the accountability duties attributed to both ATCOs and pilots, and to reduce ambiguity in the chain of command and oversight during ground operations.

More details available in B.3.3.

4.2.2.3 Professional liability – Pilot

Eventually, the legal risks for pilots are limited, recurring only in the 20% of the potentially problematic scenarios considered. However, worth to be noted that also this category of actors is subject to a qualified liability regime, according to the accountability duties related to their role. This involves they

[6] AEON (2022). D5.2 - Human Performance Assessment Report (68-71).

have comprehensive monitoring duties on the other actors involved the tactical phase of tuning (i.e., TaxiBot operators, pushback tug drivers).

More details available in B.3.4.

4.2.2.4 Corporate liability

It should be specified that the liability profiles concerning defective products, and the individual professional responsibility of personnel are, in all cases, complemented by a potential risk of corporate liability. This risk is associated with the proper implementation of the system, the adequate definition of usage procedures, the revision of current taxiing operation protocols, the retraining of personnel, and, last but not least, the correct maintenance of the systems.

The factors most likely to contribute to these risks may include, on one hand, the design of operational procedures and training programs that do not sufficiently support effective human oversight, thus limiting operators' ability to make autonomous and accountable decisions; and, on the other hand, the use of poor-quality input data, insufficient computing resources, or inadequate maintenance practices, all of which can undermine system performance and reliability.

These potential liability issues may involve the airport managing entity as a whole, as well as, more specifically, the companies employing the various operators involved, such as air carriers, ANSPs, and ground handling service providers.

More details available in B.3.5.

CRT-ASTAIR-TRL1-ERP-02.04.02

Liability impact of innovations: Identification of suitable measures in design, organisation, and policy to mitigate identified risks.

Results:

In light of the findings, and considering the current maturity level of both the solution and the associated ConOps, the following recommendations can be made:

- Ensure the highest possible level of compliance with the guidance set out by EASA for the development and deployment of AI-based solutions, with particular emphasis on AI assurance objectives and operational explainability.
- Review and refine the human–AI interaction flows, especially in scenarios involving sequential or shared responsibilities across multiple actors and AI and delegation of authority
- Minimise by design potential communication overlaps or ambiguities that could degrade operators' situational awareness and impair the effectiveness of human oversight over Ai functioning during operational execution.
- Reassess the AI levels indicated in the sequence diagrams to ensure they are fully aligned with the actual content and criticality of each task, particularly in view of explainability and human oversight requirements.

EUROPEAN PARTNERSHIP
Co-funded by the European Union

To support the safe and effective deployment of the solution as it reaches higher maturity levels, the following recommendations are advised to address potential issues related to warning defects and organisational accountability of both AI providers and deploying organisations:

- Ensure clear and complete communication of system requirements and constraints, including input data quality, computing resources, functional capabilities, known limitations, vulnerabilities, and maintenance needs, to support informed integration and operation.
- Maintain high standards for input data quality, computing infrastructure, and system maintenance practices, to preserve performance reliability and operational integrity throughout the system's lifecycle.
- Provide comprehensive user guidance covering system usage, interpretation of outputs, recognition of abnormal behaviours, and associated operational protocols, to enable effective and safe operator interaction with the system.
- Clarify expected system usage and interaction modalities, ensuring that users are fully informed of correct procedures and the potential consequences of improper use, thereby promoting operational consistency and predictability.
- Design operational procedures and training programmes that actively support human oversight, enabling operators to retain autonomous decision-making authority and accountability in line with regulatory expectations.

CRT-ASTAIR-TRL1-ERP-02.04.03

Liability impact of innovations: Positive feedback from AB stakeholders on the proposed concept or suggestions for alternative enhancements.

Results:

Although liability assessment was not a specific focus of discussion with AB stakeholders during the validation activities, the topic was introduced during the Final Workshop to gather initial insights on potential issues and concerns. The feedback received was generally positive regarding the validity of the adopted approach. Particular attention was drawn to potential risks related to product and professional liability, especially in view of a possible integration of the system into the A-CDM framework. These bottom-up insights are consistent with the top-down analysis presented earlier and have been taken into account in formulating the final recommendations.

CRT-ASTAIR-TRL1-ERP-02.04.04

Liability impact of innovations: Ensuring that the concept does not introduce unacceptable liability risks for actors and stakeholders.

Results:

Considering the expected level of maturity of ASTAIR, the ConOps does not introduce unacceptable risks for the actors involved. However, in view of future developments, the tasks assigned to the GND ATCO as well as the explored dependencies with the tasks performed by the AI and the TFM should be analysed better. The possible risks detected can be adequately mitigated by implementing the measures suggested above and could be better managed, in practice, even by a better definition of the legal or regulatory status of the TFM in the respect of the GND ATCO.

Co-funded by the European Union

4.2.3 OBJ-ASTAIR-ERP-03 Results

Objective: **Assess the operators' controlling and engaging with the automation at diverse levels.**

Title: Interaction with different automation levels

R&I Need: Operator's controlling and engaging with the automation at diverse levels

Results:

Overall, the results were OK.

The participants demonstrated a high level of engagement throughout the validation runs (for both Solution Scenarios), with minor decrease from Run 1 to Run 2 of the Solution Scenario 1. They all expressed positivity towards the system, even if sometimes they felt like having a passive role more than a proactive role, especially during routine tasks in Solution Scenario 2 Limitations were addressed.

However, customisation options might have some limitations and constraints, which have been addressed, even though it should be taken into consideration that ASTAIR project has TRL1.

CRT-ASTAIR-TRL1-ERP-03.01

Description: Level of Operator Engagement: Assess the extent to which operators actively interact with the automated system and utilize its features to enhance operational efficiency.

Results:

The level of operator engagement was assessed through the End-of-Day questionnaire, direct observations, the User Engagement Scale (UES), and final debriefing sessions conducted after each validation runs of the Solution Scenarios.

The participants expressed strong satisfaction with the ASTAIR tool and demonstrated a clear interest in understanding its underlying logic. Following an initial familiarisation session, all participants showed both the capability and willingness to engage with the system's key functionalities, including traffic integration, conflict inspection, and the delegation of tasks between human operators and the AI agent. Their active involvement reflected a positive attitude towards the tool and its potential for operational integration. Furthermore, they all expressed openness to AI and new modes of operations during the debriefing sessions for both of the Solution Scenarios

The overall level of engagement was notably high. Indeed, the results of the UES questionnaire demonstrated a constant engagement throughout the runs that decreased a bit from Run 1 to Run 2 of the Solution Scenario 1 (mean run 1= 3.25, mean run 2= 3.19). This might be because the scenarios were non-nominal and, therefore, more complex. Specifically, from the results of the UES, they didn't evaluate the system as confusing. On the contrary, they found the tool to be both valuable and aesthetically appealing. During the debriefings, operators expressed positive attitude towards the tool's usability and effectiveness, saying that the experience was rewarding. The majority indicated that they could envision themselves using such a tool in the airport operations in the future.

EUROPEAN PARTNERSHIP
Co-funded by the European Union

From the operators' perspective, they were asked to indicate whether they perceived their role as proactive, neutral, or passive while engaging with the system and its functionalities. The results revealed a range of engagement experiences, as illustrated below (Figure 7).

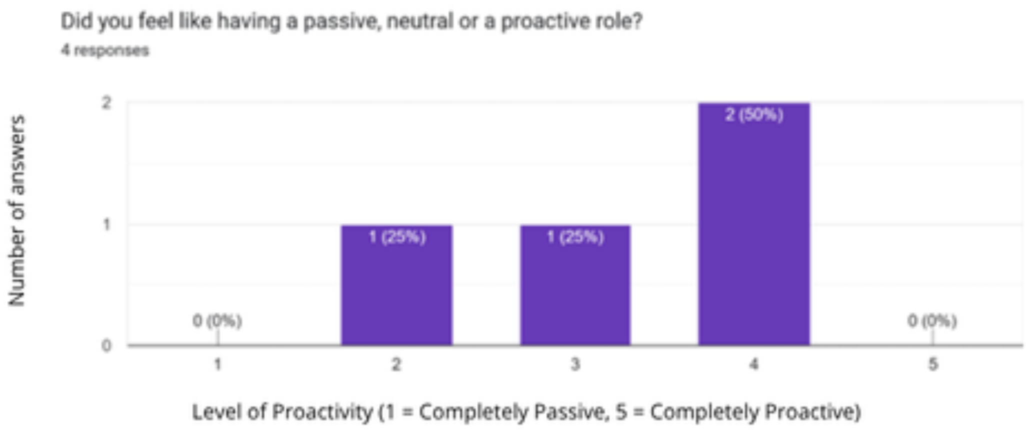

Figure 7: Answer to the question "Did you feel like having a passive, neutral or proactive role?" to the End of day questionnaire

This result was confirmed also during the debriefing sessions, in which many participants reported feeling relatively passive, particularly in the absence of conflicts to resolve, most notably during the Solution Scenario 2 runs. Indeed, they reported feeling like an observer because the AI was computing everything on its own.
One participant didn't actively interact with the tool within the first 5 minutes of the simulation and were in a supervisory/monitoring role only. Nevertheless, in the long term, most of the participants expressed a sense of control, noting that the AI's ability to compute conflict-free routes and allocate stands provided them additional time and cognitive capacity to concentrate on other tasks or address more complex conflicts and situations requiring their attention. This was shown both in Solution Scenario 1 and 2.

AI routes overriding:
At any time, the easiest way for the ATCO to override AI automatic guidance is simply by giving a clearance to the pilot directly over radio. The ATCO can also manually modify the trajectory on the radar image by simple drag and drop, this would result in a new digital clearance emission. As it is shown on the image below, it has also been anticipated that the ATCO could manipulate the trajectory with a finer grain and specific action to be directly drawn on the trajectory.

EUROPEAN PARTNERSHIP
Co-funded by the European Union

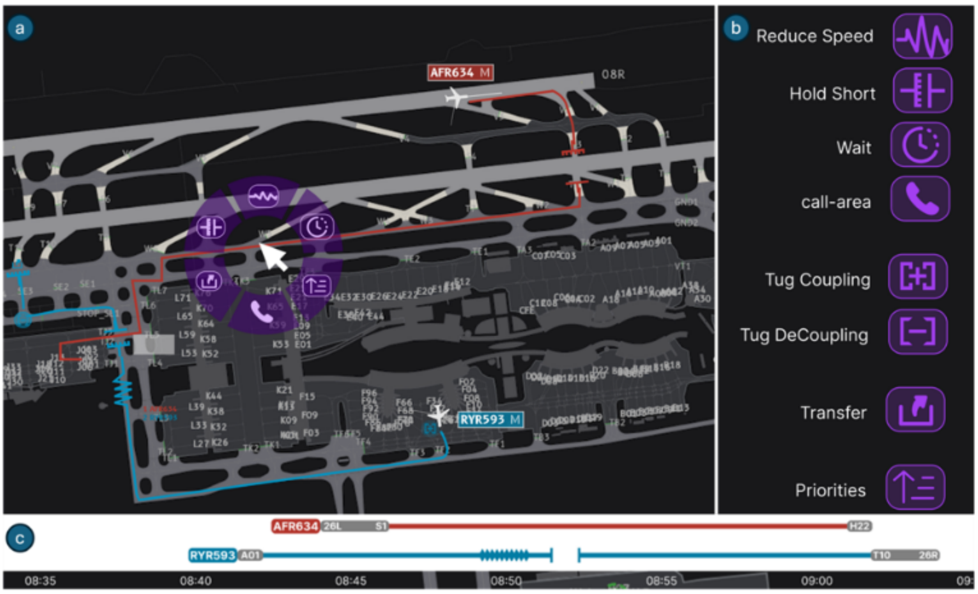

Figure 8: AI trajectory features

These features have not been practically validated in the real time validation.

Nevertheless, ATCOs confirmed that manual route edition functionalities would be useful and thus have a potential in improving the operational feasibility of the concept, while having the positive impact on the human performance, such as workload decrement.

However, two potential limitations related to user engagement should be addressed:

- Simulation Effect: A significant number of participants reported experiencing a sense of "simulation effect." This phenomenon can lead to overreliance on the tool in a simulated environment and may prevent effective transfer of skills or knowledge to real-world settings.
- Lack of Realistic Scenarios: Some participants perceived the scenarios used in the Inspection interface as unrealistic or not taking into account real operations and environments constraints. This perception could further limit the applicability and transferability of the experience to actual operational environments.

These limitations may affect the extent to which insights and learnings from the simulation can be effectively applied in real-world contexts. However, in the ASTAIR validation runs, many of these risks were proactively mitigated by involving real air traffic controllers (ATCOs), using multiple data sources (e.g., questionnaires, SAGAT, and debriefing sessions), incorporating real-world operational data and communication structures, and allowing sufficient time for participants to familiarise themselves with the system prior to testing.

CRT-ASTAIR-TRL1-ERP-03.02

Description: customisation options - Ability to effectively control and engage with the automation before and during the operation: Measure the range and effectiveness of customisation features available to operators for adjusting system behaviour and settings.

Co-funded by the European Union

Results:

The customisation options were assessed during direct observations, questionnaires and debriefing sessions.

Both during training phase and Run 1 and Run 2 of Solution Scenario 1, all participants were engaging with the automation during their operational tasks. Most of them, when asked, felt like they were in control of the situation.

As shown in the OBJ-ASTAIR-ERP-01 Results, almost all participants were able to change routes for the aircraft, and they were actively checking what the AI was doing, what were the future trajectories and possible conflicts. However, there were some customisation options that they felt were missing, especially prioritisation of conflicts, as they couldn't prioritise, based on their expertise, the conflicts showed by the AI agent. This was addressed also during the Solution Scenario 2 runs, where participants noted some lack of customisation options, such as action buttons or possibility to see affected traffic.

Limitation and recommendation:

The biggest limitation, which should be addressed even if ASTAIR is a project with a low TRL (TRL1), is that participants couldn't really override or change AI inputs because a simulated AI was used. Because of this, the decisions were already planned and so it was not exactly how the AI agent could behave. Future developments and research should take this into consideration for further improvements of the tools and concept.

4.2.4 OBJ-ASTAIR-ERP-04 Results

Objective: **Assess the HMI / interactive tools and adaptive AI algorithms supporting the operators.**

Title: Usable HMI and Interactive Tools

R&I Need: HMI/Interactive tools, adaptive AI algorithms

Results:

The results were OK.

The HMI usability was marginally acceptable, indicating that some users experienced some difficulties, which have been addressed throughout the debriefing sessions. However, the participants stated that the interface helped them improve their situational awareness and performance, which is positive for the operational feasibility of the concept. The HMI and interactive tools also enhanced and supported the decision-making of operators, providing timely information. The tools performance was addressed as OK too for TRL1.

Some recommendations and limitations were addressed for future development of the HMI, also in the light of knowing that the tools are in their early stages and, therefore, the SUS score of 62.5 is quite positive

CRT-ASTAIR-TRL1-ERP-04.01

HMI Usability: Assess the ease of use and intuitiveness of the HMI design, based on operators' ability to quickly understand and navigate the interface.

Results: HMI usability was evaluated through the post-run System Usability Scale (SUS) questionnaire, the End-of-Day questionnaire, direct observations by project consortium members and debriefing sessions conducted after each validation runs.

The System Usability Scale (SUS) score (M = 62.5, SD = 4.56) is considered slightly below average, indicating marginal or just acceptable usability. The average SUS score typically falls around 68. A score of 62.5 may be classified as “OK” or “Marginally Acceptable,” approximately corresponding to a grade of D or C on an academic grading scale from A to F. This score suggests that, while the interface is functional, users may experience some degree of difficulty or frustration during use.

In response to these findings, debriefing sessions and qualitative feedback have been conducted to gain further insights. It is important to note that the system is still in its early development stages; therefore, this score can be considered as a reasonable and positive starting point, taking into account the low maturity level, with clear opportunities for improvement.

During the direct observations, except for one participant who required additional time to become familiar with the system, all participants quickly understood the functionalities of the interface both in Solution Scenario 1 and Solution Scenario 2. However, The Inspection interface in Solution Scenario 1 was perceived as less easy to get familiar with, probably needing more training time in the future, and during the debriefing sessions, participants noted, in both Solution Scenarios, that while the HMI is clean and effective, it lacks certain critical information, which diminishes the efficiency of decision-making. The participant who experienced initial difficulties attributed this to age-related factors and a differing mindset toward adapting to new technologies. In contrast, younger participants demonstrated a more intuitive understanding of the HMI and its features, reflecting a more technology-oriented approach. This generational difference suggests that, if the system will be deployed, its acceptance and effectiveness may be positively influenced by the technological adaptability of future operators.

Furthermore, the questionnaire results indicated that participants generally found the system to be user-friendly and expressed a willingness to use it frequently. However, some participants also reported that they would require assistance from a technical expert to initially become proficient with the system (Figure 8). This suggests that, while the overall usability is high, there may be a need for additional support or onboarding for certain users during the initial adoption phase.

Co-funded by the European Union

Figure 9: Answer to item "I think that I would need the support of a technical person to be able to use this system" from SUS (System Usability Scale)

Aligned with this result, another item from the System Usability Scale (SUS) reflects a similar perspective. Specifically, the questionnaire results indicated that some participants experienced fewer difficulties than others in understanding what was required to begin using the interface (Figure 9).

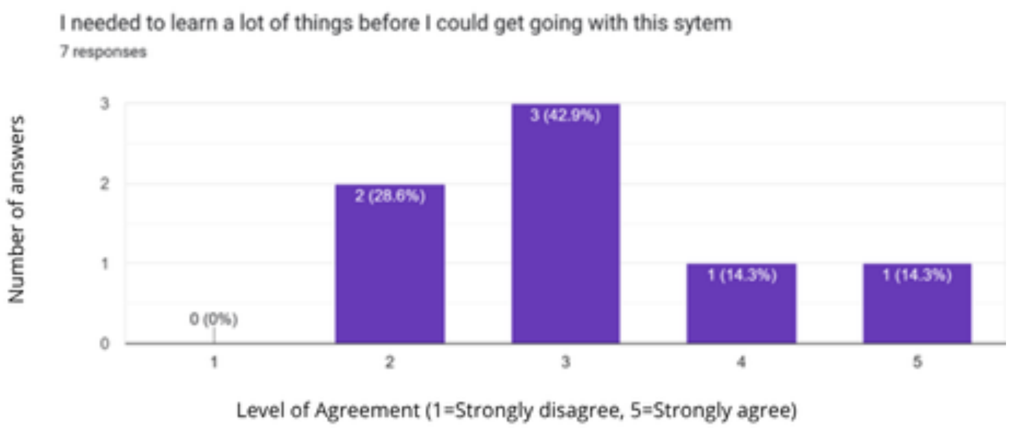

Figure 10: Answer to item "I needed to learn a lot of things before I could get going with this system" from SUS (System Usability Scale)

These findings are consistent with the insights gathered through both direct observations and debriefing sessions. A likely explanation for the observed differences in system interaction may be attributed to factors such as age and professional background. Notably, one participant was not an Air Traffic Control Controller (ATCO), which understandably contributed to the need for additional technical support during the runs.

However, most validation participants perceived the HMI as usable and provided positive feedback, especially considering that ASTAIR is TRL1 and at very early stages.

CRT-ASTAIR-TRL1-ERP-04.02

Decision-making support: Measure the effectiveness of interactive tools in providing operators with relevant information and assistance for making real-time decisions during ground operations.

Results:

Overall, all participants became more and more comfortable using the interactive tools to collaborate with the automation as they progressed through the Validation runs.

Decision-making support emerged as a key topic in both the Solution Scenario 1 and 2. Significant insights were gathered on this subject through the Situational Awareness Rating Technique (SART) assessments, SAGAT technique, and the debriefing sessions.

From the SART results, participants demonstrated a moderate level of situational awareness across both runs of Solution Scenario 1, with a slight decrease observed from Run 1 (S.A. = 19.5, SD = 4.04) to Run 2 (S.A. = 18.0, SD = 1.00). This decline may indicate an increased level of difficulty during Run 2, likely due to the introduction of non-nominal scenarios, which appear to have affected participants' overall situational awareness.

Indeed, during the debriefing session, they said that they felt that the information displayed on the interfaces were useful and not overwhelming. They also highlighted the usefulness of some features of our tools such as the hazardous situations crosscheck monitor

On the other hand, from End of Day questionnaire results, participants found the interactions timely, and they felt supported in case of conflicts (Figure 10):

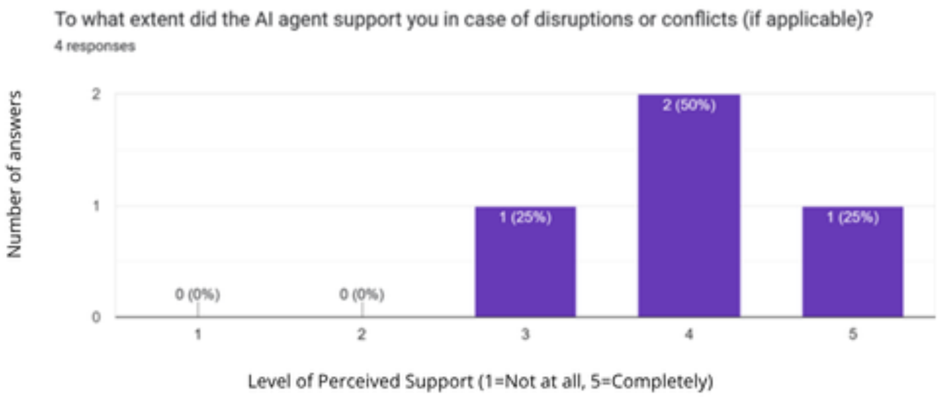

Figure 11: Answer to the question "To what extent did the AI agent support you in case of disruptions or conflicts?" from the End of day questionnaire

To additionally assess the situational awareness of the participants, a SAGAT questionnaire was performed by each participant during the final validation. In these runs (Solution Scenario 1) the participants had to manage the traffic on the Paris CDG airport platform while using ASTAIR and procedures that come with the use of tools. A question on situation awareness was asked to the participants after all the simulation screens were frozen and hidden, eight times at random during the runs. The questions covered the three levels of situation awareness: perception, comprehension and future. Answering the questions about perception required the participants to remember the information they had seen on the interfaces of the ASTAIR tools. The questions about comprehension

EUROPEAN PARTNERSHIP
Co-funded by the European Union

required the participants to recall forthcoming or occurring hazardous situations on the platform using either the A-SMGCS or the automation supervision tool. The questions about future were designed to ensure that the participants had used the automation plans Inspection interface before giving the right answer. Participants scored 1 point for any good answer and no point for any wrong answer.

The SAGAT average score was 75% (SD=12.5%). However given the sample size of our study, the SAGAT scores cannot measure the performance of the controllers with our interactive tools. The scores for each participant show that the situation awareness was different for each of the air traffic controllers that participated in the study (Figure 11). It is worth noting that P4 managed to build a good situation awareness using our tools scoring more than 87% in SAGAT.

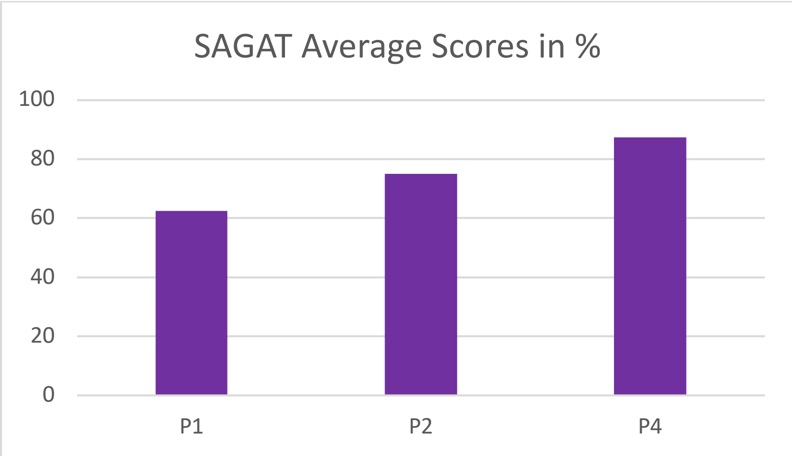

Figure 12: SAGAT average scores

Interestingly, the scores show that the participants had an excellent perception of the information using the interactive tools from Solution Scenario 1, suggesting that the user interfaces were well designed and the information displayed to the controllers were clear and unequivocal. Although the comprehension score was slightly below the perception score, it reached more than 80%. This suggests that the participants could understand most of what occurred on the taxiways during the SAGAT questionnaire and our tools could provide a good support for decision making in real time.

It seems that participants struggled to inspect what the automation's taxiing plans for upcoming arrivals and departures flights as the SAGAT future score only reached 50% (Figure 13). This score reveals that either participants did not use our Inspection tools to explore the automation plans regularly or the tool can be significantly improved. One of the participants mentioned that "he did not inspect future automation plans as he trusted the AI to provide the best plan".

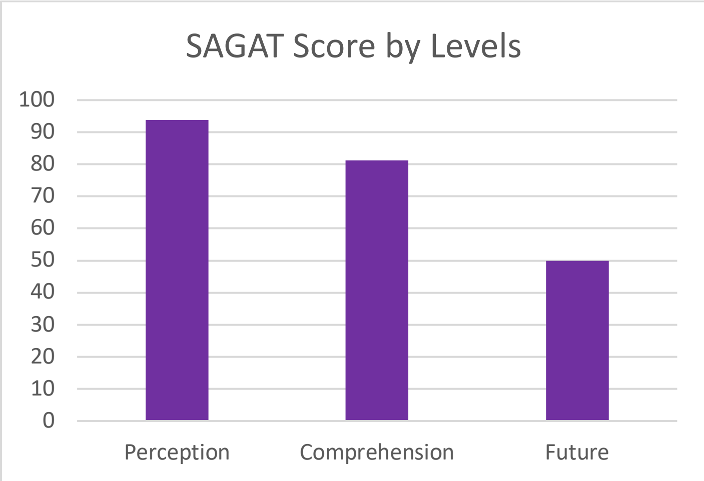

Figure 13: SAGAT score by situation awareness level

Despite the overall good perceived situational awareness, during the debriefing sessions, participants identified specific features of the HMI that may require implementation or enhancement to better support the situational awareness:

- Conflict alerts presented a challenge in both the Solution Scenario 1 and 2. Most participants failed to notice the alert of certain issues, such as arriving flights without assigned parking stands, returning flights due to medical emergencies, and various delays. During the debriefing sessions, participants consistently recommended enhancing the visibility and salience of these alerts to better capture attention.
- The prioritisation of conflicts was another key topic in the Solution Scenario 2 debriefing sessions. Most participants repeatedly reviewed the conflict resolutions, and one participant noted that, in his view, certain conflicts were of lower importance than others. However, there was no effective means, either automated or manual, to prioritise them accordingly. Because of this, the users would prefer to have a way to prioritise the conflicts accordingly.
- Some participants reported uncertainty regarding whether conflicts had been resolved, indicating a lack of feedback, particularly from the pilot's side. As a result, some of them continued to use radio communication to alert pilots, as they were unsure whether the pilots were aware of the situation or had received the necessary information. This issue was also raised during the Final Workshop of the project. One stakeholder emphasized that pilots should receive up-to-date information and maintain situational awareness through the system, including clear guidance on ongoing events and required actions.
- The Inspection interface was sometimes dismissed during the validation runs. While participants demonstrated an understanding of how to use it, some of them chose not to engage with it. This was primarily due to the added cognitive load it introduced and its lack of user-friendliness. Many participants expressed a preference for having this feature either on a separate screen or integrated as a layer within the primary interface.
- Regarding the Solution Scenario 2, participants noted the absence of several key pieces of information, including certain alerts from MAS (e.g., stand occupancy), directional labels,

delays, Target Take-Off Times (TTOT), and traffic impact. They indicated that the inclusion of such information would enhance the tool's ability to support efficient decision-making.

From a situational awareness standpoint, participants generally found the system's changes to be clear, but the implications of those changes were less clear (Figure 13). Many did not use the Inspection interface, which was intended to clarify the AI agent's planned responses to hazardous situations. This might happen because, as mentioned by one of the ATCO, the Inspection interface was in another screen and this was not too convenient to be used, since it requires more physical motion and switching from the radar image

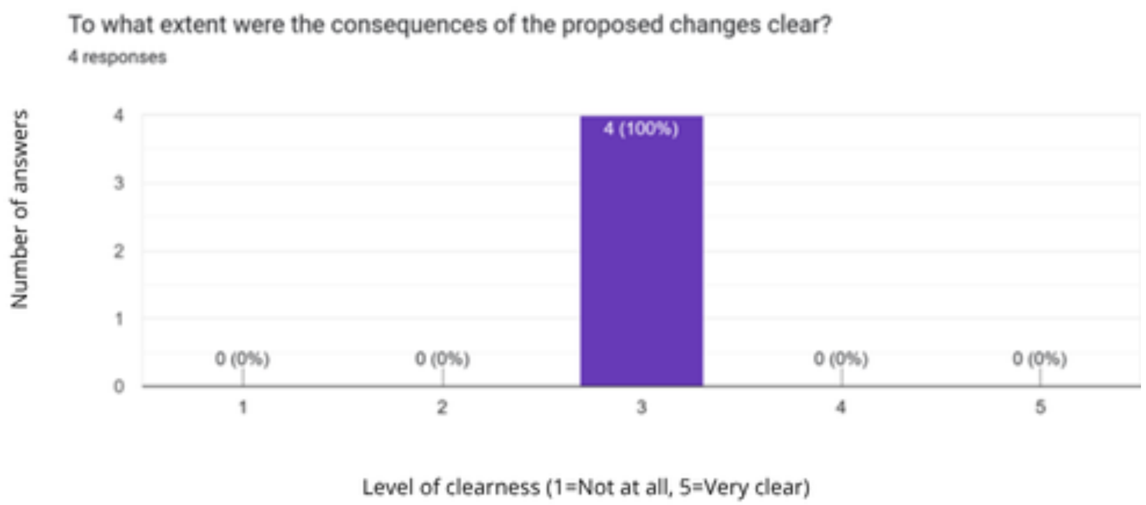

Figure 14: Answer to the question "To what extent were the consequences of the proposed changes clear?" from the End of day questionnaire

However, the system successfully maintained a high level of participant engagement throughout the validation runs. Participants reported feeling alert and prepared for the tasks (Figure 14), which is particularly significant given the demands of Air Traffic Controller (ATCO) activities.

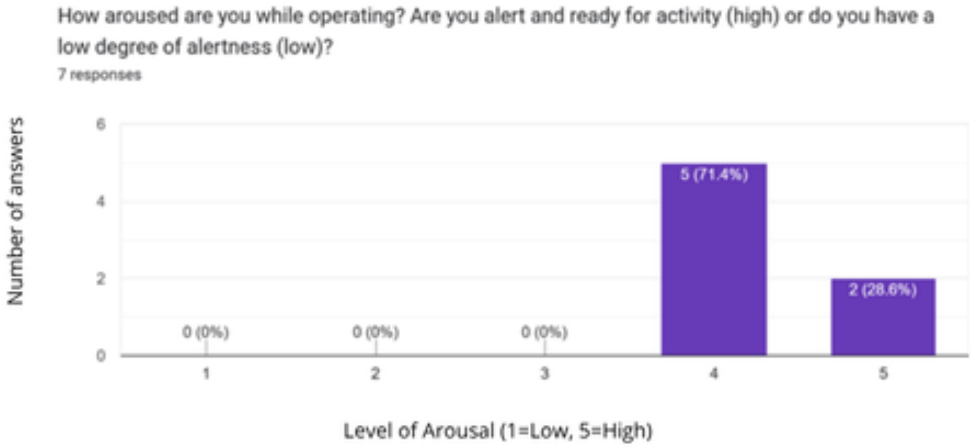

Figure 15: Answer to the item "How aroused are you while operating?" from SART (Situational Awareness Rating Technique)

CRT-ASTAIR-TRL1-ERP-04.03

Tools performance: Evaluate the tools performance and the impact on the efficiency of operators' interactions with the HMI and interactive tools.

Results:

From the perspective of tool performance, participants consistently reported, during debriefing sessions, that the AI demonstrated strong capabilities, particularly in supporting traffic integration and conflict inspection. These strengths were especially evident in Solution Scenario 2, where the AI effectively managed complex operational demands and contributed to maintaining situational awareness and workflow efficiency.

While participants did not frequently intervene manually, demonstrating a general confidence in the AI's decision-making and the system's logic, they did occasionally adjust assigned routes or slots (Figure 15). This limited intervention was also due, in part, to the fact that participants could not conveniently make changes because a simulated AI has been used, restricting real-time control. As a result, the responses from the End-of-day questionnaire reflected a balanced perception of trust in automation, combined with an ongoing need for human oversight and flexibility in specific operational contexts.

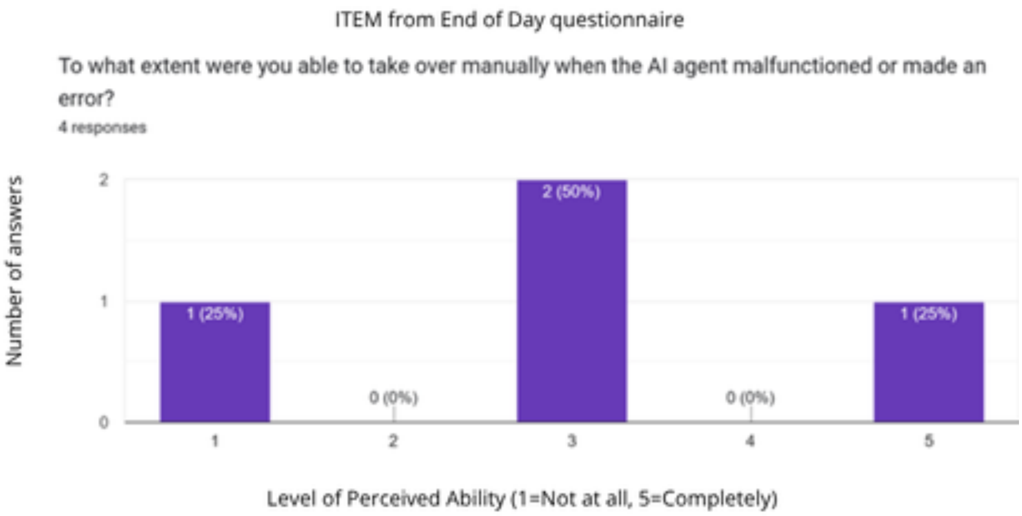

Figure 16: Answer to the question "To what extent were you able to take over manually when the AI agent malfunctioned or made an error?" from the End of day questionnaire

Regarding the Solution Scenario 2, participants expressed a high level of trust in the AI. Indeed, all of them indicated that they would not have taken actions different from those recommended by the MAS. This indicates a strong level of performance and reliability demonstrated by the tool. However, participants noted that they would have appreciated the ability to make certain adjustments, such as prioritising specific aircraft, accounting for additional delays, and taking into consideration communication with pilots.

Recommendation:

A key recommendation emerging from the validation runs is the need for greater transparency in the rationale behind AI decisions. While participants generally trusted the AI's outputs and found its conflict detection and resolution capabilities effective, they expressed a desire to better understand the reasoning behind specific actions or suggestions. This was particularly evident in situations involving complex traffic interactions, where operators wanted reassurance that the AI's decisions were based on sound logic.

4.2.5 OBJ-ASTAIR-ERP-05 Results

Objective: **Assess the Optimized Path & Motion Planning for Efficient Ground Operations.**

Title: Optimised Path & Motion Planning

R&I Need: Path Planning

Results:

The results were OK.

Fast-Time Simulations confirmed that the ASTAIR system maintains or increases airport capacity and improves taxi time performance, supporting compliance with CTOT slots and optimised runway usage. Conflict-free routing was ensured through accurate path planning, respecting operational safety requirements and providing safe trajectories. Tug resource management was improved, with simulations showing reduced fuel consumption of outbound taxiing aircraft as ETV fleet size increased, up to the identified optimisation point.

CRT-ASTAIR-TRL1-ERP-05.01

Ensure capacity is maintained or increased with the new ASTAIR concept.

Results:

Fast-time simulations (FTS) were conducted using the multiagent system (MAS) model for path planning outlined in D2.1. Two of the busiest operational days at Amsterdam Airport Schiphol (EHAM) to date were simulated. Based on the conducted fast-time simulations, the ASTAIR concept ensures that the capacity and runway throughput are maintained or increased. Participants in the validation scenarios consistently highlighted that the MAS model managed the traffic efficiently, indicating that there should be no impact on capacity.

Vehicles shall be able to follow a routing clearance with speed profile. Even though several solutions exist (TaxiBot, autonomous follow-me cars, auto-taxi aircraft), the resilience of the Multi Agent System (MAS) towards deviation from plan shall be evaluated as well as the impact on airport capacity. Measuring the tolerance margins that the MAS is able to cope with will allow to estimate the precision required for the vehicles in terms of position and speed guidance without impacting the airport capacity.

The respective results are presented in detail in the following.

Traffic overview of conducted fast-time simulations

In the FTS, two scenarios are considered based on the historical flight schedules of two specific days. Each scenario is defined by the primary Runway Mode of Operations (RMO) in use on that day: RMO North for 17th July 2019, and RMO South for 18th July 2019. Together, RMO North and RMO South represent the most common runway configurations at EHAM, providing a solid basis for evaluating standard operational conditions. Both selected days feature busy flight schedules with high volumes of taxi movements. A detailed overview of the number of flights per scenario is provided in Table 16.

(a) Traffic data

date	17-07-2019	18-07-2019
flights	1489	1492
arrivals	745	744
departures	744	748
RMO	RMO North	RMO South
RMO phases	19	19

(b) ICAO-types. WTC = wake turbulence category

	parameters		count per day	
	shape [m]	WTC	17-07-2019	18-07-2019
ICAO-A	12	CAT-F	0	0
ICAO-B	25	CAT-E	22	20
ICAO-C	40	CAT-D	1195	1198
ICAO-D	54	CAT-C	37	43
ICAO-E	72	CAT-B	213	206
ICAO-F	80	CAT-A	22	25

Table 16: FTS simulations: overview of (a) traffic data and (b) parameters and daily counts of ICAO-types

To give an overview of the traffic situation, Figure 16 shows the hourly count of all flights for both the historic (black dotted line) as well as simulated operations (grey line) over the two days of 17th and 18th July 2019. The two curves almost match each other, with the simulated operations showing a slightly lower total count due to the lower taxi times as discussed below. Furthermore, the chart visualizes the count of arriving vs. departing flights: the alternating trend between landings and take-offs that is characteristic for a hub-and-spoke airport such as Schiphol is clearly visible. This is also reflected in the frequently changing RMO phases over the course of the two days, illustrated by the coloured shades in Figure 16.

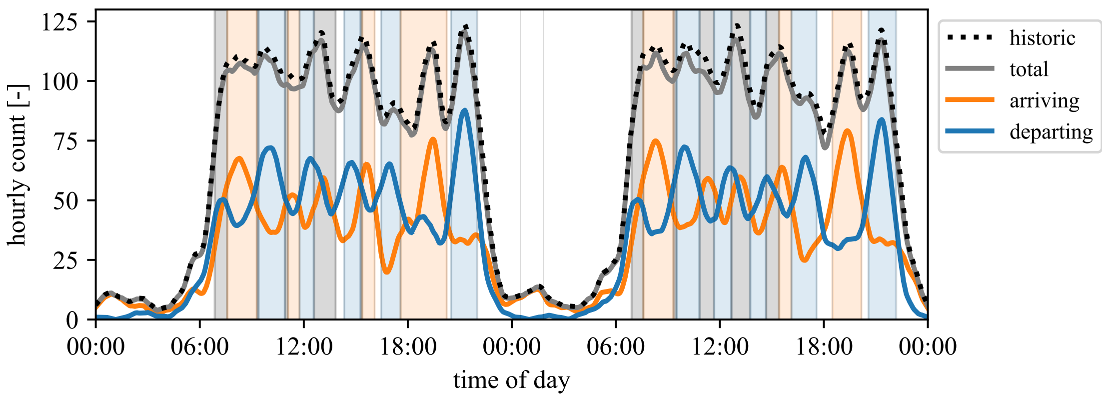

Figure 17: Hourly count of flights over the two days. Shades denote the RMO phase: off-peak (white), arrival-peak (orange), transition (grey), and departure-peak (blue)

Summary of performance indicators

Table 17 lists the mean, median, and interquartile range (IQR) of the taxi times for inbound and outbound flights. The underlying distributions exclude flights that hold explicitly during taxiing, i.e. arriving flights without a free gate and departing flights with an issued CTOT-slot exceeding the required taxi time. The distributions are further discussed in the following section, while the excluded flights are analysed separately in another section. As predictability metrics, list the prediction error $RMSE$ and inequality coefficient U, and refer to these in the section on predictability below. Moreover, for any runway in use, the maximal throughput and occupancy rate per hour have been reported. For arrivals, the indicators are identical between the historic and simulated operations as the actual landing time (ALDT) has been used as spawn-time in the simulation. For departures, the maximal hourly throughput is similar between historic and simulated operations, while the maximal hourly occupancy rate has increased for the simulated operations. More details on the runway sequence and capacity as well as the listed CTOT-slot violations are further discussed in own sections below.

Date		17-07-2019		18-07-2019	
Operations		Historic	simulated	historic	simulated
ARR	mean taxi time	04:38	03:03	10:15	07:40
	median taxi time	04:00	02:53	10:44	08:37
	IQR taxi time	02:51	01:40	05:59	05:13
	RMSE taxi time prediction	01:47	00:02	02:16	00:02
	U taxi time prediction	33.9%	1.1%	20.6%	0.5%

	RWY throughput*	42		38	
	RWY occupancy*	68.7%		69.3%	
DEP	mean taxi time	14:29	11:18	10:20	07:05
	median taxi time	14:32	11:20	10:06	06:34
	IQR taxi time	06:32	04:56	04:20	02:29
	RMSE taxi time prediction	02:45	00:40	02:44	00:35
	U taxi time prediction	18.1%	5.5%	25.2%	7.9%
	RWY throughput*	45	45	43	42
	RWY occupancy*	74.7%	76.1%	73.5%	77.0%

*: maximal hourly value for any runway
**: corrected after checking historic A-CDM milestones

Table 17: Comparison of historic and simulated operations with $w_{plng} = 20\text{min}, h_{plng} = 10\text{min}$

Comparison of taxi time distributions per runway

Figure 17 displays the distributions of taxi times for each arrival and departure runway for the historic and different simulated operations as box-and-whisker plots. These represent the median, first and third quartiles as box, while outliers are marked by whiskers and points. Like Table 17, all distributions exclude flights that hold explicitly during taxiing, which are analysed separately in a following section.

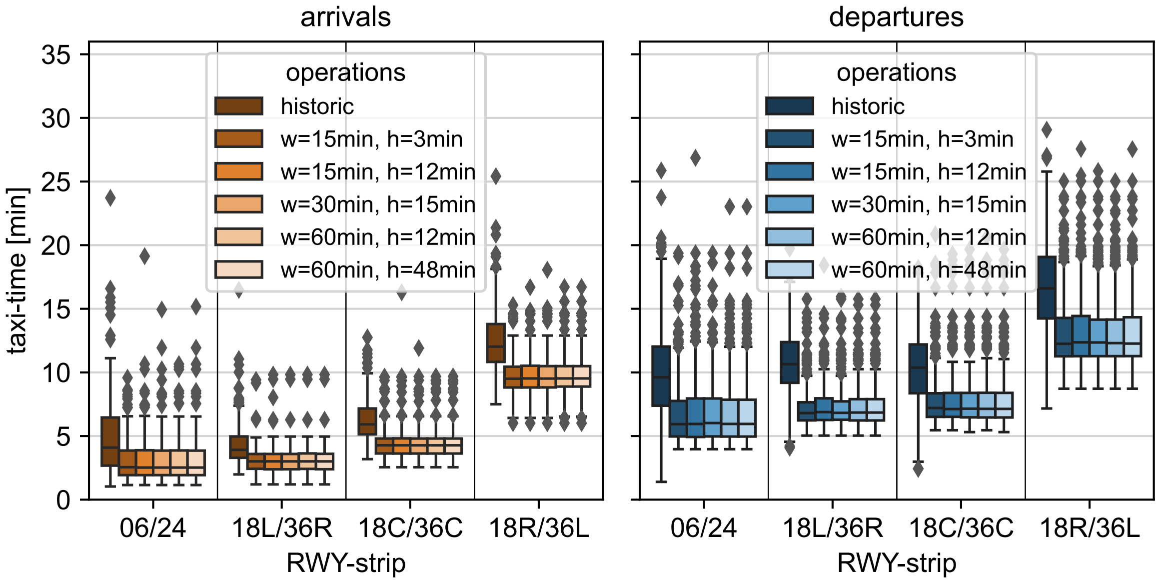

Figure 18: Box-and-whisker plot of historic and simulated taxi times for arrivals and departures per runway-strip

In general, the taxi times from the simulated operations are shorter and vary less. Since the runway 18R/36L is far away from the central part of Schiphol, taxiing to/from this runway takes more time than to any of the other runways. As the departing aircraft have to start their engines to taxi after pushback, their taxi time to any runway is in general longer than for aircraft that land on the same runway. Furthermore, since an engine-start time of 6 min has been used for large aircraft (ICAO-D to

ICAO-F) in comparison to 3 min for small aircraft, the taxi times of departing aircraft vary more than those of arriving aircraft in the simulated operations.

Of the simulations carried out for the algorithmic analysis, the five shown in in the figure represent different combinations for w_{plng} and h_{plng}. Values have been tested for w_{plng} below 15 min, but the routing algorithm did not succeed to find a solution throughout each of the two days. The taxi times do not differ significantly between the simulations, both for arriving and departing flights from all runways. Therefore, the conclusion is that both the planning window w_{plng} as well as the replanning period h_{plng} do not impact the efficiency of operations within the ranges that have been tested.

Predictability of taxi times

Figure 18 shows the variability of the taxi time predictions with respect to the remaining actual taxi time for four sets of w_{plng} and h_{plng}. The red lines mark the 1 % and 99 % quantiles as indication of the accuracy over the remaining time. For all four simulations, the first predictions underestimate the actual taxi time: the aircraft start taxiing almost at the end of the planning window and most conflicts are thus not yet resolved. The deviation to the actual taxi time decreases strongly in the following planning rounds. When the remaining taxi time is less than w_{plng}, the difference between predicted and actual taxi time is negligible for more than 50 % of all flights. The accuracy further increases towards the end of taxiing. Longer planning windows yield accurate predictions within a longer duration till the end of taxiing, supporting Hypothesis H2, while h_{plng} has a subordinate effect on the predictability. As listed in Table 17, the $RMSE$ and U values decrease significantly in comparison to the historic operations. Note that this may change when deviations to the planning arise during execution, which has not been modelled in this work.

Holding of inbound and outbound flights

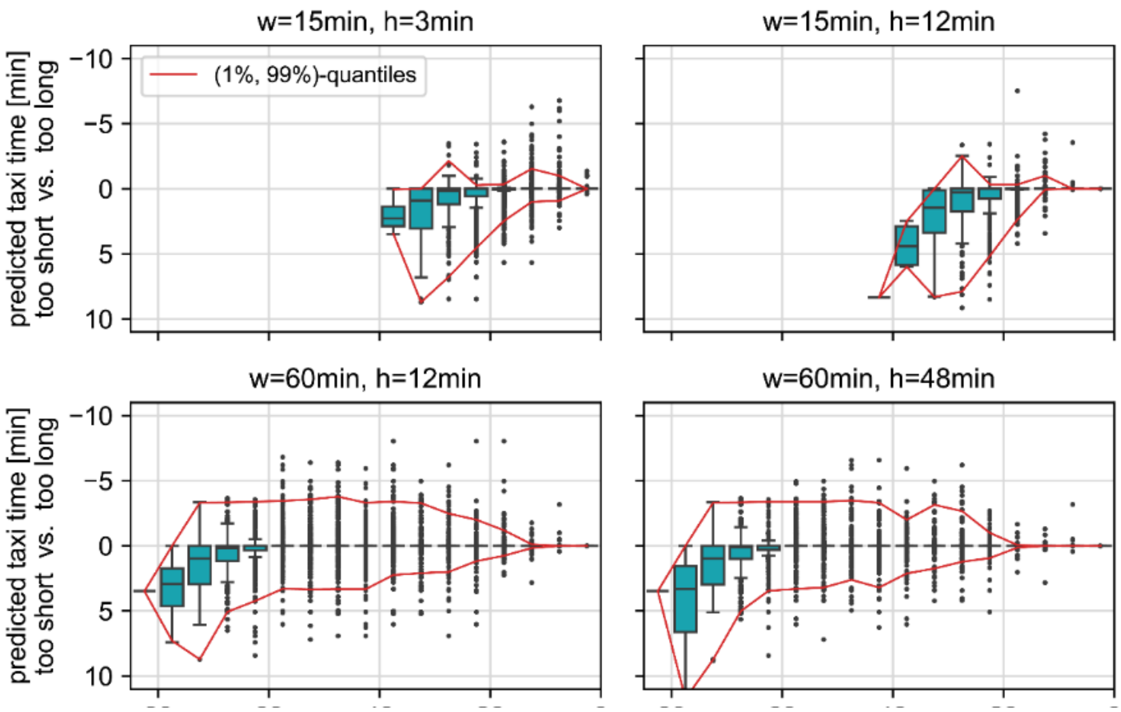

Figure 19: predictability of taxi times

EUROPEAN PARTNERSHIP

Co-funded by the European Union

In the taxi time analysis above, flights with a hold-type assigned by the Routing Agent have been excluded. Figure 19 compares the taxi times of these flights between historic operations and the different hold-types of the simulated operations as box-and-whisker plot. In general, not many flights are holding. In comparison to Figure 17, the taxi times of flights with inbound holding are significantly higher, and are similar between historic and simulated operations, also considering that some of the historic A-CDM milestones end at the holding locations. For most outbound aircraft that must be delayed due to their CTOT-slots, the Routing Agent lets them hold at their stand. The taxi times of the

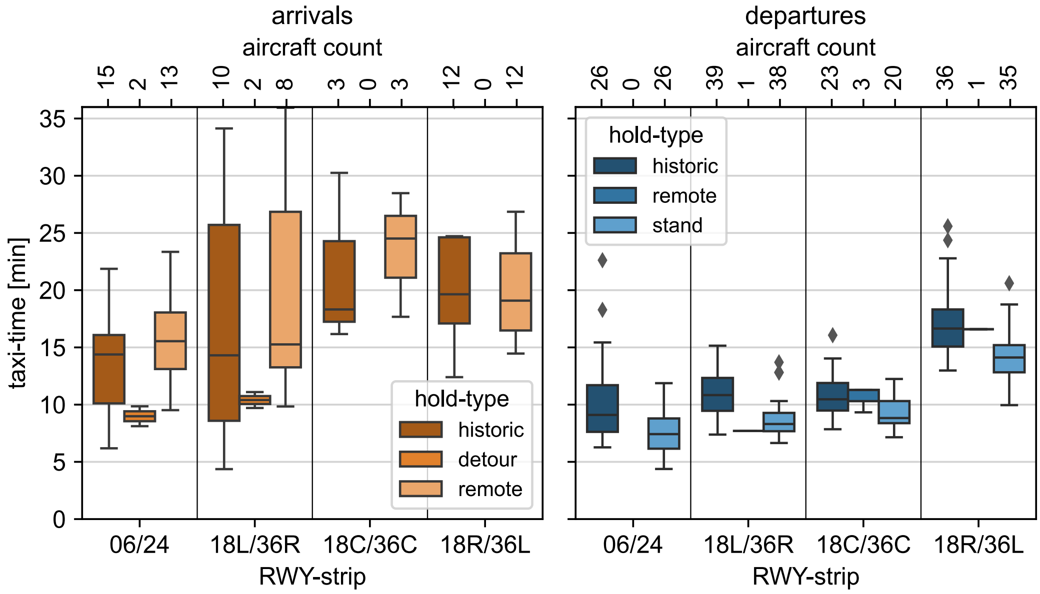

Figure 20: Box-and-whisker plot of historic and simulated taxi times dependent on the hold-type of arrivals and departures per runway-strip; simulation with w_plng=20min,h_plng=10min.

historic operations are slightly longer, which has an influence on the moment that the aircraft take off within the CTOT-slot, as has been analysed further below.

CTOT-slots of outbound flights

Over the two days, a total number of 442 CTOT-slots are assigned by Eurocontrol. Figure 20 visualizes the compliance between the take-off times to the CTOT-slots of both historic and simulated operations. While the historic times almost follow a normal distribution centred around the time issued by Eurocontrol, those of the simulated operations are skewed towards the beginning of the CTOT-window. In its current implementation, the routing algorithm optimizes for lowest taxi times and does not attempt to let aircraft take off closer to their calculated take-off time. From the 442 flights, 8 historic flights (1.8 %) do not comply with their CTOT-slots. In comparison, only a single simulated flight takes off outside its CTOT-slot. However, it has been noted that for this flight, the historic A-CDM milestone occurs multiple minutes after the actual pushback, rendering it a faulty outlier.

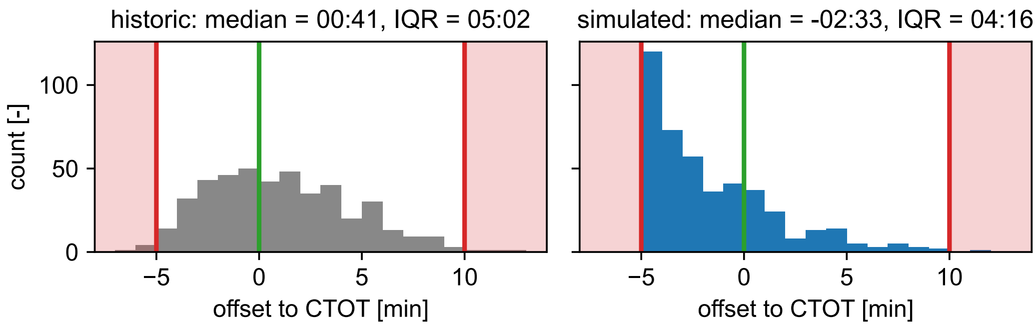

Figure 21: Compliance to CTOT-slot (white area) of historic (left) and simulated operations (right)

Runway sequence and capacity of outbound flights

In the fast-time simulations, pre-defined take-off sequence has not been used. Thus, the order of aircraft departing from a runway is different between the historic and simulated operations, as exemplary shown in Figure 21. The minimal separation between flights according to RECAT-EU is illustrated by the red shades. While the historic order mostly adheres to the RECAT-EU separation, in some cases, two aircraft are separated less than the minimum. It has been confirmed with the track data that indeed some flights take off around 52s after the previous one, despite having a larger separation than indicated in the figure due to inaccuracies of the A-CDM milestones.

As emergent property of the MAS, whenever possible, flights are grouped together with minimal

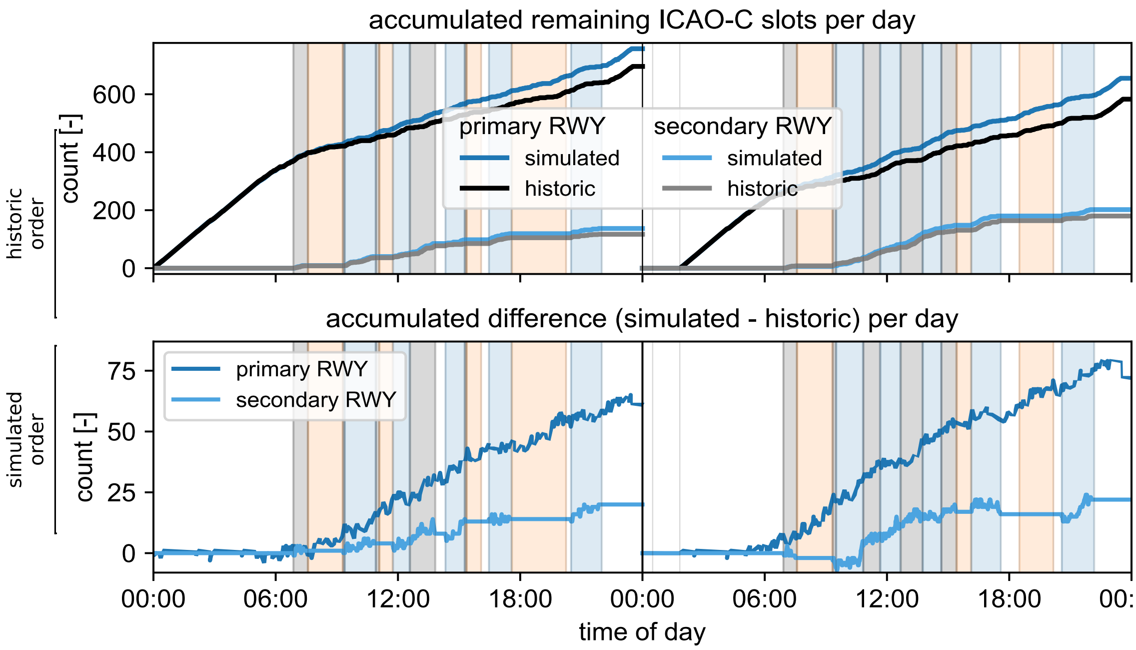

Figure 22: Comparison of exemplary take-off order at runway 36L between historic (top) and simulated operations (bottom), with actual take-off time (blue lines) and minimal WTC separation (red shades)

separation time between each take-off. Based on this observation, the remaining slots have been counted per departure runway for each of the two days. For both historic and simulated operations, take-off slots remain even during busy departure peaks marked by the blue shades. However, throughout the two days, more slots remain for the simulated operations: in relation to the total number of departing flights, 10.9 % and 12.6 % additional slots are available on each of the two days, respectively. This suggests that the MAS better utilizes the potential runway capacity.

CRT-ASTAIR-TRL1-ERP-05.02

Conflict-free routing: Evaluate the safety implications of optimized taxi routes, including collision avoidance measures and adherence to operational regulations and guidelines. Conflict-free routing (Conflict detection & resolution) to ensure safety levels remained.

Results:

As outlined in D2.1 [5], the MAS model accounts for aircraft shapes and kinematics during path planning on a high-resolution airport layout and was calibrated using historic ADS-B data. Furthermore, important airport surface movement elements and processes were explicitly included in the model such as pushback, engine-start, inbound holding, complying with CTOT-slots, and adhering to a minimal safety distance during taxiing as well as minimal wake turbulence separation during take-off.

Based on the conducted fast-time simulations above, the routing algorithm was evaluated to be well suited for planning conflict-free trajectories of all vehicles moving on the airport surface. No conflicts occurred in any of the fast-time simulations. This indicates that the ASTAIR concept maintains the required safety levels.

CRT-ASTAIR-TRL1-ERP-05.04

Maximise the usage of tugs / tugs resource management.

Results:

Fast-time simulations (FTS) were conducted using the tug fleet management (TFM) algorithms outlined in D2.1 [5], and based on the same traffic data at EHAM as provided in CRT-ASTAIR-TRL1-ERP-05.01. The conducted FTS showcased that the ASTAIR concept maximises the use of tugs. The respective results are presented in the following.

Analysing the Effect of ETV Fleet Size

In this section, the impact of fleet size on total fuel consumption is analysed. The conducted simulations used a scheduling interval of 40min and a horizon of 60min.

Figure 22 shows the relation between the number of ETVs and the total fuel consumption of outbound taxiing traffic for the two scenarios. The dashed blue line indicates the baseline fuel consumption; in the case all regional and narrow-body aircraft are towed to the runway.

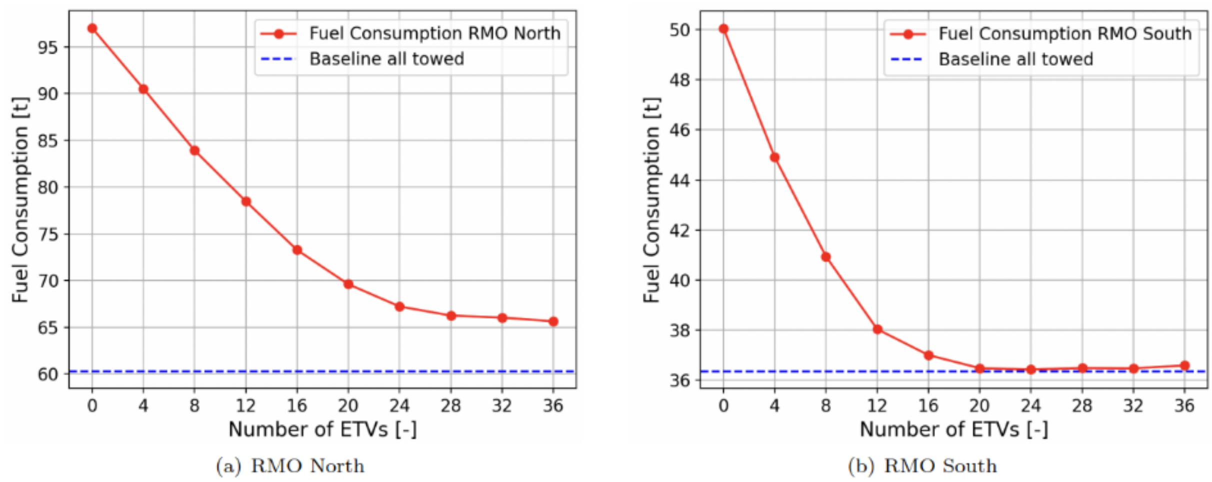

Figure 23: Effect of ETV fleet size on total fuel consumption of outbound taxiing aircraft for RMO North and RMO South. Results averaged from n = 5 runs. Error bars indicating standard error of the mean omitted for clarity due to their negligible size.

From the fuel consumption curves in Figure 22, it can be concluded that the TFM algorithm effectively utilizes an increasing number of available ETVs to decrease the total fuel consumption of taxiing aircraft. However, there is a distinct point of diminishing returns which is strongly dependent on the RMO. Overall, ETV operations have the potential to reduce the total fuel consumption of outbound taxiing aircraft by 38% for RMO North, and 28% for RMO South. In absolute terms, this results in a daily reduction of 111 and 44 tonnes of CO2 for the two scenarios, respectively.

The difference between the two scenarios can be explained by the runway configurations of the two RMOs. In RMO North, the preferred runways, 36L and 36C, involve relatively long taxi times. While this increases the potential for fuel consumption reduction, it also necessitates deploying more ETVs to achieve these savings. Conversely, in RMO South, the preferred runways 18L and 24 are closer, allowing ETVs to complete more tasks. However, the shorter taxi times also limit the potential for fuel savings.

When examining the relation between ETV fleet size and fuel consumption for RMO North, a clear limitation of the algorithm becomes evident. Although there are enough ETVs available to carry out all tow tasks, the algorithm struggles to produce a schedule that approaches the fuel consumption of the baseline scenario, where all eligible aircraft are towed to the runway.

The reason for this is the return route for ETVs after decoupling at runway 36C. After decoupling, ETVs must go around the entire runway, along 18C, to arrive back at the apron. This means the time to the next task is very long, leading to a low heuristic value (see D2.1 [5]). Instead, decoupling at either the Romeo or Juliet TRP has a much shorter return time with only a slight increase in fuel consumption, leading to more favourable heuristics. While this heuristic design allows the TFM algorithm to find efficient schedules with a limited ETV fleet, it steers the algorithm away from the optimal solution when the number of ETVs is practically unlimited. This behaviour is mostly expected, as the TFM algorithm was designed and optimized to handle a limited fleet.

For RMO South, increasing the number of ETVs beyond 20 has a negligible effect on the total fuel consumption. However, with only 12 ETVs, around 90% of the potential fuel savings are already realized.

4.3 Confidence in validation results

4.3.1 Limitations of validation results

The ASTAIR validation results support the achievement of TRL1 and provide valuable insights into the feasibility and potential benefits of introducing automation in airport ground operations. However, a number of limitations constrain the generalisation of the results to broader SESAR sub-operating environments and higher maturity levels:

- **Limited number of participants**: Due to scheduling and availability constraints, the number of Air Traffic Controllers (ATCOs) and domain experts participating in the validation exercises, especially the RTS, was limited. As such, the validity of some human performance results is reduced.

- **Algorithms not demonstrated in integration with the main HMI interface**: The core algorithms (e.g. the MAS routing and TFM algorithms) were validated independently, particularly through Validation Scenario 2. While they were not demonstrated in integration with the main HMI during the Validation Scenario 1(Adaptative Automation Level) in the RTS because of their computational complexity. This affects the completeness of user interaction evaluations, and the realism of the human-AI collaboration tested during the real-time simulations.

- **Simplified technical environment**: While realistic data and scenarios were used, the HITL RTS setup used a simulated AI (a scripted AI with predefined behaviours), which limited the ability to evaluate dynamic adaptation, real-time updates, or user overrides of AI behaviour. This reduces the representativeness of certain results, especially in high-tempo or unexpected situations.

- **Simulation Effect:** A significant number of participants reported experiencing a sense of the "simulation effect." This phenomenon can lead to overreliance on the tool in a simulated environment and may hinder effective transfer of skills or knowledge to real-world settings.

- **Lack of Realistic Scenarios**: Some participants perceived the scenarios used in the Solution Scenario 2, which uses Automation level 3, as unrealistic. This perception could further limit the applicability and transferability of the experience to actual operational environments.

- **Limited coverage of non-nominal scenarios**: Although one non-nominal case (e.g. sick passenger requiring rerouting) was tested, the scope of such scenarios was narrow and did not cover a full spectrum of degraded or emergency conditions. This limits insights into system resilience and operator re-engagement in failure scenarios.

- **Assumptions affecting applicability**: Some key assumptions, such as the presence of fully interoperable A-CDM and A-SMGCS systems, availability of datalink communications, and actor compliance with AI routing, may not hold at all European airports. This restricts the extrapolation of findings without further adaptation or infrastructure investment.

- **Incomplete validation of certain objectives**: Specifically, validation objective **OBJ-ASTAIR-ERP-05.03**, assessing the manoeuvrability of tugs and tug-aircraft combinations via FTS, was not addressed due to resource unavailability. As such, this aspect remains unvalidated and should be reconsidered in future R&I phases.

- Questionnaires, such as Post-Run Questionnaire and End of Day questionnaire, due to the non-integration of MAS algorithms, have been completed by participants only after the Solution Scenario 1 and they were not being used for Solution Scenario 2.

Overall, the validation confirms the potential of the ASTAIR concept operational feasibility under controlled, nominal conditions and chosen non-nominal use cases. However, these results should be interpreted with caution when considering broader deployment or extrapolation to different operational environments, particularly those with limited automation enablers (medium-small airports) or significant variability in procedures and infrastructure.

The above-mentioned limitations do not have a negative effect on the maturity of the concept due to the maturity being very low – TRL1.

4.3.1.1 Quality of validation results

The validation results are based on both qualitative and quantitative data. These include structured observations, questionnaires (e.g. SUS, UES, SART, SAGAT), debriefings, and system interaction logs. Fast-Time Simulations (FTS) supported technical assessments such as conflict-free routing and tug fleet management.

Each RTS run included a short approx. 15 min. debriefing session. The overall quality and confidence in the results is mainly influenced by **limited sample size**. A small number of participants, particularly ATCOs, restricts the generalisability of findings to different airport environments and complexities.

Conclusion: The quality of the results is assessed as **medium**. The findings support the achievement of TRL1, but further validation with more participants, higher tool maturity, and realistic AI integration is needed in future phases/technology readiness levels.

4.3.1.2 Significance of validation results

The results have a high significance due to the various stakeholders involved in the validation activities (RTS, workshop), ranging from specialist with ATC background to the airport operations specialists who have detailed understanding on the ground movements processes and procedures.

Moreover, different validation and results assessment techniques were used, including both qualitative and quantitative data analysis, thanks to the various questionnaires and debrief sessions, and quantitative FTS results.

Lastly, numerous runs were executed during the RTS week, which covered a wide range of ASTAIR Use Cases.

Below is a summary of operational significance of the results. Operational significance here concerns operational realism of the different validation scenarios which depends on a number of factors which are very much dependent on the chosen environment.

- **Simulated AI**: In some scenarios, scripted AI was used instead of a real-time adaptive system, limiting interaction (some ASTAIR functionalities, such as “slow down”) and realism of the operational feasibility, and realism of the workload, human-AI collaboration, etc. Some functions were present on the HMI, but they were not implemented and thus not used. For instance, the Future inspection action buttons (slow down, ...).

- **Simplified environment**:

    - The validation did not fully replicate A-SMGCS environments (and its all functional blocks):
    - Surveillance Service – simulated,
    - Routing Service: core of ASTAIR (ASTAIR extends AEON's multi-agent routing for conflict-free trajectories with speed profiles). Integrates with A-CDM to adjust TSATs and arrival parking assignments.
    - Guidance Service: Not implemented, only referenced. Visual guidance elements (e.g. stop bars, TCL) were not part of the validation setup. ASTAIR assumes future integration for speed regulation compliance.
    - Safety Support / Alerting Service: Partially addressed through concept. No full CATC, CMAC, or RMCA implementation. The AI system detects potentially unsafe trajectory intersections (e.g. “close calls”) and displays them for the ATCO.
    - Operational communication workflows: No live R/T voice communication between actors (e.g. ATCOs and pilots) in Validation Scenario 2.
    - Some roles, like the Tug Fleet Manager, as defined in the concept Use Cases were not fully integrated into the one ASTAIR system and thus the TFM role and functionality were evaluated in a separate session with a ground movement expert.

- **Tools low maturity**: The ASTAIR tools are at an early development stage – TRL1. Some usability issues were noted (e.g. SUS score = 62.5), and features such as conflict prioritisation were not fully implemented.

- **Training needs**: The participants were lacking the Schiphol airport environment and operations knowledge as they were experienced on the CDG. This was relevant for Scenario 2.

Despite these limitations, the results confirm user trust, system usability, and technical feasibility. FTS confirmed the routing system’s ability to maintain safety and capacity.

5 Conclusions and recommendations

5.1 Conclusions

This section provides a summary of conclusions derived at a solution level regarding operational feasibility, technical feasibility, performance, as well as conclusions on the solution maturity.

5.1.1 Conclusions on project/ SESAR solution maturity

The ASTAIR project initiated its activities at Technology Readiness Level 0 (TRL0), with the objective to investigate the fundamental principles, initial feasibility, and conceptual feasibility of introducing automation into airport ground operations. Following the successful execution of the planned exploratory research activities, the ASTAIR solution has achieved Technology Readiness Level 1 (TRL1), in accordance with the SESAR 3 Joint Undertaking maturity framework for exploratory research.

Throughout the three validation exercises conducted during the project lifecycle, a structured and iterative assessment approach was applied, combining expert workshops, prototype evaluations, Human-in-the-Loop Real-Time Simulation, Fast-Time Simulation, and stakeholder feedback collection. This approach allowed for a thorough evaluation of the operational and technical feasibility and initial acceptability of the ASTAIR Solution across its key research areas, including operational feasibility, human-AI collaboration, HMI usability, adaptive AI algorithms, and motion planning capabilities.

The validation results provided initial evidence that the ASTAIR concept is feasible and relevant for the intended operational environment, within large complexity European airports using A-CDM and A-SMGCS. The concept shows initial benefits in terms of improved predictability, operational efficiency, sustainability and safety of ground operations. Human-AI collaboration models demonstrated their applicability for supporting supervision and coordination of airport surface movements.

Several limitations remain, as expected at this maturity stage. The integration of the algorithmic modules with the main HMI was not fully demonstrated during validation, due to technical and scheduling constraints. Certain stakeholders, in particular pilots and airlines, were not fully involved in the RTS. Further developments are necessary to address these aspects, including improvements in AI transparency, task allocation between human operators and automation, and clarification of procedural, legal, and liability aspects for future maturity levels.

Based on the concept developed and evidence collected, the ASTAIR solution fulfils the criteria for TRL1-achieved maturity level. The exploratory research activities have successfully confirmed the basic principles of the concept and identified the key areas requiring further investigation and development.

5.1.2 Conclusions on concept clarification

Overall, the concept that has been validated during the real time exercise is really close to description given in D1.3 Initial Concept Outline [4]. However, it was initially proposed to modify the radio communication workflow between pilots and ATC to facilitate future trajectories integration:

EUROPEAN PARTNERSHIP
Co-funded by the European Union

- For a departure flight, the pilot would contact ground ATC a few minutes before TSAT to confirm it. The rationale is to give valid inputs to the multi agent system that computes the conflict free trajectories.
- In the same manner, for an arrival flight the pilot would contact ground ATC to confirm the runway exit and timing.

Although it could make sense for a departure flight, and it could help the ground ATCO deal with a departure flight a bit in advance and at the very last moment, this workflow for an arrival flight is not applicable since it would require the pilot to call ground ATCO while still in tower ATCO sector. It must be noted that no apron control position was set up during the run.

In conclusion, these information confirmations should be done electronically, and the current radio communication workflow should not be impacted by ASTAIR solution.

In comparison the to the concept described in the initial concept, an additional functionality has been implemented. Indeed, to facilitate the human operator understanding of the trajectories computed by the Multi agent routing system, an interpretation layer has been added. It consists first of an analysis of each vehicle trajectory to detect anormal speed up or down to detect regulations applied for deconfliction and in a second step, all trajectories are analysed altogether to detect potentially conflicting situations. Actually, even though all trajectories computed by the multi agent system are deconflicted via speed control in theory, it could happen that some crossings lead to close-call, several aircraft coming in the same intersection in a short period of time. These situations, once detected, are displayed on the supervision HMI that allows future trajectories integration and on the real time radar image as a reminder to the supervisor to check that there is no safety issue. This feature has been implemented on trajectory provided by the multi agent system, which are the actual trajectories used by the vehicles, but it could be applied to trajectories predictions computed on actual systems.

5.1.3 Conclusions on technical aspects

On a technical point of view, the simulation setup was considered convincing by the different participants. The A-SMGCS prototype, although not fully functional obviously (no safety nets for instance) is representative enough according to ATCO. In addition, the different scenarios proposed were realistic in terms of activities and workload on Roissy-CDG and Amsterdam-Schiphol airports.

On a side note, the Tug Fleet Manager position has not been integrated in the final setup due to time constraint. However, ASTAIR solution could be extended to the fleet management, hence it could be discussed with airport management experts on it.

It must also be noted that the setup was slightly different depending on target airports:

- On Roissy-CDG, forged data have been used to emulate a conflict-free automated traffic, and operational events were added live by pseudo-pilots.
- On Amsterdam-Schiphol a different and lighter HMI has been used to replay recorded results of actual AI (the Multi Agent System).

These different setups were after all clearly explained and adapted to the different validation scenarios.

On an operational feasibility point of view, the enablers listed for ASTAIR solution are confirmed and with the prerequisites, no major blocker has been raised for an implementation of the solution. As a reminder, the enablers required for ASTAIR solution are:

- A-CDM, departure sequencer and arrival manager to get reliable information on TSAT (target startup times) and ETA (Estimated Time of Arrival)
- A-SMGCS with surveillance and routing services
- Datalink / CPDLC communications with aircraft and tugs. Communication speed is important; it shall be fast and reliable. In order to be efficient, ASTAIR concept depends on the possibility to have reliable digital communications with the controlled vehicles. Indeed, the clearances computed by ASTAIR AI need to be sent electronically to the vehicle.
- Aircraft and tugs that can follow quite closely the trajectories and speed profiles computed for them. This is a key enabler to the concept that the vehicles are able to follow a routing clearance with speed profile. Even though several solutions exist (TaxiBot, autonomous follow-me cars, auto-taxi aircraft), the resilience of the Multi Agent System (MAS) towards deviation from plan shall be evaluated as well as the impact on airport capacity. Measuring the tolerance margins that the MAS is able to cope with will allow to estimate the precision required for the vehicles in terms of position and speed guidance without impacting the airport capacity.

5.1.4 Conclusions on performance assessments

Safety

Reduction of radio communication use facilitates important communications and relieve ATCO to focus more on traffic supervision for more safety related tasks (such as separation monitoring, clearances, etc.).

Capacity

Based on the conducted fast-time simulations, no impact on the airport capacity of the ASTAIR concept was observed. In the simulations, the high traffic demand was maintained, and the outcomes suggest that the operations are more efficient and predictable, as well as result in better use of the potential runway capacity.

Human Performance

Based on the validation results, the interactions and tools supported the ATCOs activities and supported collaboration with AI at different level. However, there are some recommendations that need to be addressed:

- Current situation and future trajectories presentation could be better integrated: validation of Solution Scenario 2 showed that the 2 screens set up may not be optimal
- Manual edition of routes suggested by the AI has not been thoroughly integrated and tested,
- The use of cross-checks facilitated the surveillance activities, but more efforts are required to create different levels of criticality and support ATCOs filtering/creating or removing some as well as fine tuning how the AI detects them.
- Non nominal UC: departure A/C with sick passenger that needs to go back to parking (UC8) creates transitional situation where some a/c are manually take out of the automated loop by

ATCO to handle the operational event -> the run stopped a few minutes after the event but it was observed that the ATCO need to be accompanied to go back to a full automation mode. However, the AI routing system using actual traffic rules allows the current full manual mode to be a safe fall back in non-nominal situations.

Liability

The ASTAIR concept carries moderate liability risks mainly around AI reliability and task delegation between humans and AI. Key recommendations include improving AI transparency, clarifying roles, and strengthening procedures and training to ensure safe and accountable operations.

5.2 Recommendations

5.2.1 Recommendations for next R&I phase

- **Validation**
    - **Key Technologies**: Testing the feasibility of proposed autonomous taxiing solutions (e.g., tugs, onboard systems) under operational conditions. This includes validating ecological routing algorithms and speed regulation systems.
    - **Other end users / effected stakeholders' involvement in the validations (RTS and workshops)**, such as pilots, tug fleet manager, airlines, tower controller.

- **Integration with Existing Systems**: Continuing the development of integration between ASTAIR and current airport systems such as A-CDM and A-SMGCS to ensure seamless operations. Better integration of inspecting tools into existing A-SMGCS.

- **Human-Automation Teaming**:
    - Exploring advanced collaboration between human operators and automated systems to ensure safety and manage complex airport environments.
    - Handover when automation fails, i.e. degraded automation (keep lower level of automation (L1A-L2A) + recommendation tools)

- Implementation of the above recommendations for the next phase into the concept development, validation and reporting process.
- Further evaluation of **operational feasibility for other stakeholders** concerned by the solution is needed: airport operator / vehicle drivers (winter service, firefighting service, wildlife service, airfield inspection drivers etc.). The change management of the solution for different operational services need to be conducted.
- **Pilots feedback** has been addressed through workshops, but they didn't participate to the RTS activities of Validation Exercise 03. Further evaluation of operational feasibility for pilots is needed for next phases.
- **Operational and Safety Enhancements**: Refining conflict resolution mechanisms and safety nets for ground controllers to handle non-nominal situations effectively.

According to the vision outlined in the ATM Master Plan [21] and the EASA AI Roadmap [20], human operators in air traffic systems will increasingly delegate tasks to automated systems, creating a collaborative human-machine team. Thus, the next research should evaluate:

- The effects of **technological advancements**, particularly changes in automation levels and the shift from executive to supervisory roles, on the nature and frequency of operator interventions, required skills, and overall performance, including the impact on fatigue.
- Potential **safety risks/hazards** associated with the evolving role of human operators, such as the cognitive demands of supervisory tasks and how they might affect performance and safety.

5.2.2 Recommendations for future R&I activities

The future R&I steps following this work would typically involve:

- **HP and Safety assessments:** including Human Performance and Safety Scoping and Change Assessment activities and identification of HP issues and benefits and safety hazards.
- Generative AI for Proactive Emergency Forecasting: to develop generative AI models to simulate complex emergency scenarios (e.g., technical failures) and enable operators to test possible strategies before the emergency occurs, while feeding information to all required stakeholders (ATC, ground handling, pilots, etc.).

Liability

Promote a clearer definition of the ATCO's monitoring tasks, especially when the ConOps assumes that they should receive inputs that have no immediate impact on their sequence diagram.

Ensure better human oversight, including through explainable operations (by design and by default) and a transparency policy involving instructions, manuals and training.

"Provide a clear definition of the conditions under which authority is delegated to or from the AI for each actor involved and assess the resulting impact of such delegation on the subsequent forms of Human-AI Interaction (HAI) with other actors interacting with the system.

Promote a clearer definition of the TFM operative profile, bearing in mind the margin of action defined by Delegated Regulation (EU) 20/2025.

EUROPEAN PARTNERSHIP
Co-funded by the European Union

6 References

6.1 Applicable documents

SESAR solution pack

[1] D1.4 ASTAIR Concept Outline 01.00

[2] SESAR ASTAIR D1.2 Workshops Report

[3] D5.1 ASTAIR DES HE SESAR ERP_v.01.02, November 2024

[4] D1.3 Initial Concept Outline, v.01.01

[5] D2.1 - Support algorithms for automated tug assignment and path planning, v.01.00, May 2025

Content integration

[6] Common Taxonomy Description 1.0

[7] SESAR ATM Lexicon

Content development

[8] SESAR Operational Concept Document (OCD 2023)....

Performance management

[9] PJ19 Content Integration, D4.4 SESAR 2020 Performance Framework.

Validation

[10] DES HE requirements and validation / demonstration guidelines (3.0).

Safety

[11] Expanded Safety Reference Material (E-SRM) Core Document, Edition 1.2, November 2023

Human performance

[12] SESAR Human Performance Assessment Process TRL0-TRL8

Environment assessment

[13] SESAR, Environment Reference Material, alias: “Environmental impact assessment as part of the global SESAR validation”, 2014.

[14] ICAO Doc. 10031, "Guidance on Environmental Assessment of Proposed Air Traffic Management Operational Changes." 2014.

Security

[15] 16.06.02 D103 SESAR Security Ref Material Level.

Programme management

[16] 101114684 ASTAIR Grant Agreement, 31/05/2023

[17] ASTAIR DOA PartB-final - v2

[18] SESAR 3 JU Project Handbook – Programme Execution Framework, 11/04/2022, 01.00

6.2 Reference documents

[19] EUROCAE ED-78A Guidelines for Approval of the Provision and Use of Air Traffic Services supported by Data Communications, December 2020.

[20] EASA Artificial Intelligence Roadmap 2.0, Available online: https://www.easa.europa.eu/en/document-library/general-publications/easa-artificial-intelligence-roadmap-20

[21] SESAR ATM Masterplan. Available online: https://www.sesarju.eu/masterplan

Co-funded by the European Union

Appendix A Validation exercise #03 report

A.1 Summary of the validation exercise #03 plan

A.1.1 Validation exercise description and scope

This third (and last) ASTAIR validation exercise, TVAL.**03.0**-ASTAIR-TRL1, was validated throughout the following three validation activities whereas each using a different validation technique:

- Human-in-the-loop Real Time Simulation:
- End-users interacted with the ASTAIR's solution in a realistic environment. In this activity, more detailed quantitative and qualitative results will be gathered. During the test campaign, all the data were gathered using a mixed approach using quantitative and qualitative methodologies (questionnaires, observations, structured-interviews, debriefing, etc.), also including Human Performance evaluation tools addressing HMI usability, user workload and situational awareness. The data gathered were analysed using standardised research practice to ensure data reliability.
- Solution Scenario 1 and 2.
- Two validation platforms used for the two above mentioned scenarios.
- Fast time simulations (FTS):
- validated the technical side of the ASTAIR concept, e.g., MAS algorithm.
- Addressed OBJ05
- Final Workshop: gathered feedback from the ASTAIR end-users mainly on the concept's operational and technical feasibility, safety and human performance, and in general for the EPCs. This workshop focused on gathering the feedback mainly towards the OBJ-ASTAIR-ERP-01 - the operational feasibility of the ASTAIR concept. It provided the next steps for exploring the concept in the next maturity level, i.e., the outputs should be fed into the project final report including in this ERR.

The output of this final validation exercise allows the project to check the results against the validation objectives to assess the output and the TRL of the ASTAIR's solution.

The scope of this exercise includes:

- Simulation of real-time engine-off taxiing scenarios at one of the target airports: Paris-Charles de Gaulle and Amsterdam Schiphol.
- Assessment of AI support for ground operations, particularly for the allocation of TaxiBots, routing, and remote holding procedures.

Focus on interactions between AI, human operators (e.g., air traffic controllers, ground handlers), and automated systems (e.g., TaxiBots) during normal and disrupted operations.

A.1.2 Summary of validation exercise #03 validation objectives and success criteria

Refer to 4.1. As this report reports only on the exercise #03, the exercise validation objectives are the same as in the main part of the document, i.e. same as the ASTAIR Solution level objectives and their success criteria.

A.1.3 Summary of validation exercise #03 validation scenarios

This sub-section provides a summary of the reference and solution scenarios implemented in the scope of third exercise, and within its three validation activities (RTS, FTS and Final Workshop).

Reference Scenario:

The reference scenario is considered the current operations and tools, as described in the Concept Outline, as the stakeholder's assessed the ASTAIR scope and use cases against their knowledge and experience.

Solution Scenario 1:

A system and the HMI developed by ENAC, using Adaptive Automation Level approach.

Roles and responsibilities: The ground ATCO was in charge of apron and ground control; the tower controller was simulated.

Adaptive Automation Level: the tools and relevant procedures and information flows varied in automation level depending on the situation, as described in the Use Cases in Concept Outline document. The automation levels varied from 1A to 3A.

Controller Working Position (CWP) set up: The working position was composed of a 180° out of the window view, completed with two 27 inches screens displaying on the left the A-SMGCS radar image and on the right the HMIs dedicated to AI supervision (Figure 24).

Figure 24: Real time validation physical setup

Communication: Even though most of the clearances were digitally sent via Datalink, the controller still had a radio communication with the pilots for first contact and emergency messages if needed.

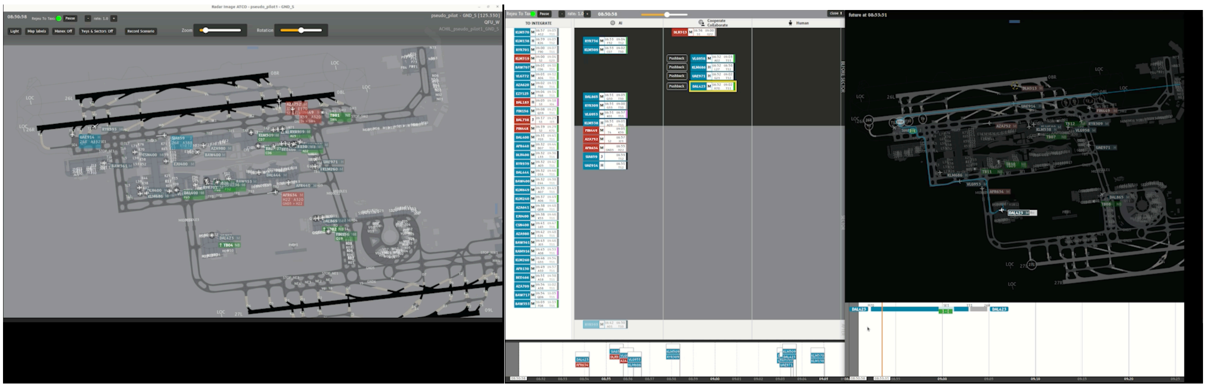

Figure 25: Close up on HMIs

Use Cases addressed:

Run 1 has addressed normal operations (UC1a and UC1b), while Run 2 has addressed non nominal scenarios (UC3 and UC8).

Normal operations: this use case is divided into two sub-use cases which are Departure with TaxiBot (1a) and Arrival with TaxiBot (1b), as addressed in the Concept Outline. Both shows the AI computing conflict-free routes and sending clearances and speed profiles, while the ATCO monitors the situation. It doesn't take into consideration delays or constraints.

Arrival aircraft with occupied parking (UC3): an aircraft is arriving, but the parking slot assigned it is not available. Therefore, the AI sends an alert to the ATCO, which will have to consider different alternatives and choose the best one.

Departing aircraft with technical issues (UC8): an aircraft departed from the assigned runway but needs to come back because of a medical emergency (passenger's sickness). The AI will alert the ATCO, and he/she will choose the best option to quickly solve the situation.

Solution Scenario 2:

A system and the HMI developed by TUD, using Level 3 automation and MAS algorithm.

Roles and responsibilities: The ground ATCO was in charge of monitoring the traffic on apron and ground control. Tower control was simulated. The traffic was managed primarily by the AI system, with the possibility for the ground ATCO to issue commands to change, adapt, or override the trajectories provided by the AI system.

EUROPEAN PARTNERSHIP
Co-funded by the European Union

Controller Working Position (CWP) setup: The working position was composed of one screen with radar image with AI decision-support Inspection interface overlay, as shown in Figure 26.

Automation and algorithm: The routing of all aircraft and towing vehicles is done based on a

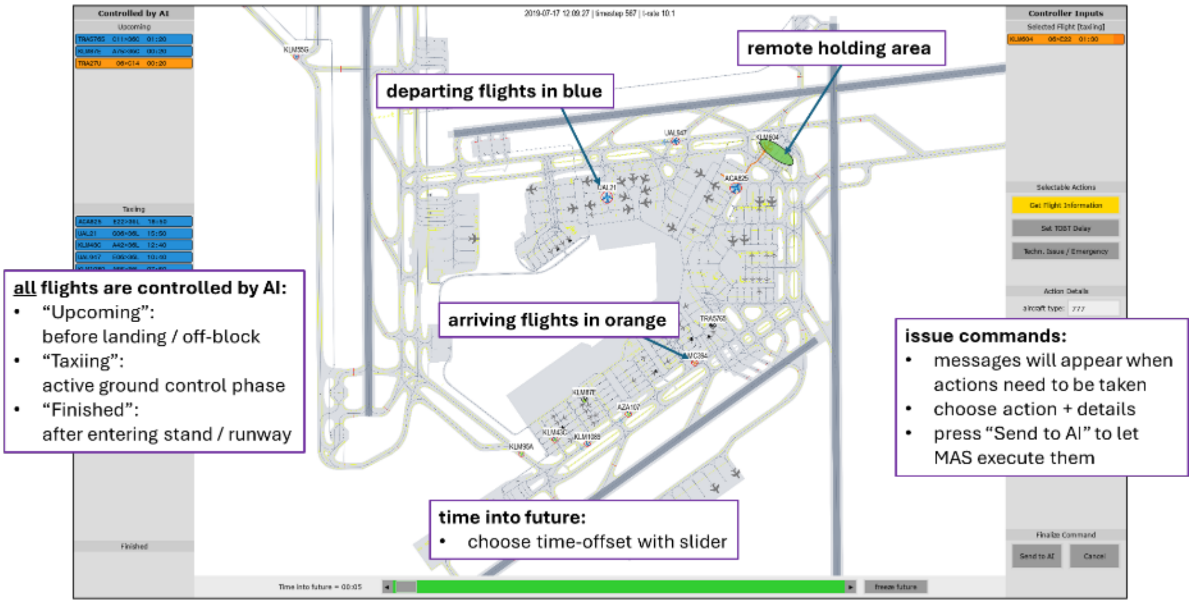

Figure 26: Setup of CWP with one screen comprising the radar image and AI decision-support overlay, summarized by the explanation-boxes in the figure.

multiagent motion planning algorithm, described in the deliverable D2.1 'Support algorithms for automated tug assignment and path planning' [5]. The two-level routing algorithm uses a low-level search to calculate individual trajectories per aircraft and coordinates all agents in its high-level search to yield conflict-free trajectories. The routing algorithm uses a rolling-horizon planning: conflicts are only deconflicted within a pre-defined time horizon. Re-planning is triggered latest after a pre-defined duration has passed, or when necessary, e.g. to adapt to input provided by the ATCO.

Use Cases addressed:

- Normal operations (UC1): the use case is divided into two sub-use cases, which are Departure with TaxiBot (UC1a) and Arrival with TaxiBot (UC1b).
- Departing aircraft with technical issue (UC8): an aircraft is coming back because of a technical emergency, and the AI sends an alert to the ATCO that needs to solve the situation.
- Normal operations with re-scheduling (UC2): the AI sends information about different delays that need to be addressed and gives support with new conflict-free routes.

Validation Objectives addressed:

OBJ02, OBJ03, OBJ04

Solution Scenario 3 - FTS:

Validation Objectives addressed: OBJ05

To be noted: during the Final Workshop of ASTAIR the use case “Automation Failure” (UC5) has been addressed with stakeholders on topics such as procedures, safety issues, human performance benefits and issues. It describes an event in which the AI stops working because of automation failure and the human operator needs to take over. However, the use case was not validated during FTS scenario.

A.1.4 Summary of validation exercise #03 validation assumptions

Validation assumptions relevant to this exercise are found in Chapter 3.3.1.

A.2 Deviation from the planned activities

Deviations from the planned activities are reported in Chapter 3.4.2.

A.3 Validation exercise #03 results

A.3.1 Unexpected behaviours/results

A minor issue has raised during Validation Exercise 03, in Solution Scenario 1. Pseudo-pilots required more time to answer than actual pilots. No other unexpected behaviours or results to declare.

A.3.2 Confidence in results of validation exercise #01

Confidence in validation results is reported in 4.3.

A.3.3 Quality of validation exercises results

Quality of validation exercise results is reported in 4.3.1.1.

A.4 Conclusions

Conclusions towards the project results and including on the third exercise are reported in Chapter 5.1

A.5 Recommendations

Recommendations are reported in Chapter 5.2.

Appendix B Liability Assessment Report

B.1 Introduction

This part of the document reports the results of the application of the Legal Case methodology to the ASTAIR solution. In particular, one of the main benefits of the early application of this method is the possibility to solve liability risks and problems with mitigations introduced at the level of operational concept, when this is still quite flexible and modulable at this stage.

In this regard, the purpose of the Legal Case application at the earliest stages of the design process is twofold.

- On the one side, this approach facilitates the detection of possible liability risks and problems of liability allocation among the different actors involved in the overall process of design, development, testing, training and operational usage of the new operational concept and associated tool, that may affect their acceptability within the organisation.
- On the other side, the Legal Case application allows the identification of suitable mitigation measures to be adopted to reduce such risks and problems.

The application of the Legal Case methodology to ASTAIR builds upon the outcomes of the human factors analysis and is informed by the descriptions of the initial ConOps (ASTAIR, D1.3), the UCs addressed during the validation activities, and their corresponding sequence diagrams. More specifically, the analysis focused on the following operational scenarios:

- UC1 – Normal Operations: Departure (UC1.a) and arrival (UC1.b) with TaxiBot
- UC2 – Normal Operations with Rescheduling
- UC3 – Arriving Traffic with Occupied Parking
- UC4 – High-Level Taxi Strategy Tuning
- UC5 – Automation Failure
- UC8 – Arriving Flight with Technical Issue

The scope of the assessment is limited to the actors and scenarios directly involved in these use cases, particularly focusing on technology providers, Ground ATCO, TFM, and the Pilot-in-Command (PIC).

The report is structured in three main sections. Following a brief introduction to the applied methodology (B.2), it presents the regulatory framework governing taxiing operations (B.2.3). After that, there is an overview of the relevant liability regimes (B.2.4), including a dedicated focus on the emerging liability issues in relation to AI in aviation (B.2.4.3). The final part provides a liability risk analysis for each of the actors considered (B.3). The general findings are presented in the section of this report dedicated to the validation objective concerning liability (5.1.4).

B.2 Methodology – The Legal Case

The Legal Case is a methodology with an associated tool intended to support the integration of automated technologies (including AI) into complex organisations, particularly in ATM. Its purpose is to address liability issues arising from the interaction between humans and automated tools, ensuring that these issues are clearly identified and dealt with at the right stage in the design, development, and deployment process.

B.2.1 Purpose and scope of the method

The Legal Case (Contissa et al., 2013) can be applied to any ATM concept involving automation, i.e., the use of automated technology, including AI. In this framework, automated technology means any "device or system that accomplishes (partially or fully) a function that was previously carried out (partially or fully) by a human operator". Thus, the notion of automation is not limited to "full automation", where an entire task is completely delegated to a machine, but rather covers cases where humans and machines interact, with machines supporting the human operator and, in some cases, enhancing and augmenting their capabilities.

The Legal Case has been designed to be flexibly applied across all the phases of maturity in a system's life cycle. The methodology can be applied both proactively (from V1 to V3 of European Operational Concept Validation Methodology - E-OCVM) and retroactively (from V4 on, of E-OCVM). Depending on the maturity phase of the technology, the Legal Case analysis will rely on different types of background information, can be used for different purposes, and will provide different sorts of output.

The Legal Case can be approached both proactively and retroactively. In the proactive scenario, the liability assessment helps support and enhance the design phase of a new operational concept or system, addressing potential legal issues that may arise from future accidents or malfunctions. In the retroactive scenario, it can be applied to existing technologies to evaluate their inherent or contextual legal risks, which may evolve over time in response to changes in the surrounding environment.

It is worth noticing that in none of these cases the Legal Case is intended to apportion liability and blame people or the organization, conversely it is intended to enforce the safety culture of the organisation making all the actors involved aware of the liability risks associated with their roles, tasks and activities and proactively identify suitable mitigations.

The method in fact entails the "design according to liabilities" approach, according to which liability is to be considered one of the inherent properties of the socio-technical system, in the same way as safety, human performance, security and environmental sustainability, and as such shall be taken into account since the earliest phases of an operational concept design.

B.2.2 The process

The Legal Case method offers a structured approach and process for the identification, analysis and mitigation of liability attribution issues related to the introduction of new operational concepts and tools in complex environments, in particular ATM.

The Legal Case process consists of the following four steps:

- **Understand context and concept**. This step involves collecting and elaborating background information about the object of the study so as to understand its socio-technical and normative aspects. The information collected concerns the operational concept itself, the context of its deployment, and the legal and regulatory aspects. This step includes the identification of the AI level of the concerned system, its impact on roles, tasks and responsibilities and a set of use cases considered relevant for the following legal analysis.
- **Identify liability issues**. This step involves identifying the possible liabilities related to the object of the study and determining the associated liability risks.

- **Address the liability allocation**. This step involves analysing the acceptability of liability risks for all stakeholders, proposing also possible mitigations that may improve liability allocation, and making design recommendations accordingly.
- **Collecting findings and Systemic Analysis**. This step presents the results of the study, highlighting the liability issues associated with the object of study and the ways to deal with legal risks, as well as making further recommendations.

The diagram below shows the workflow of the Legal Case method.

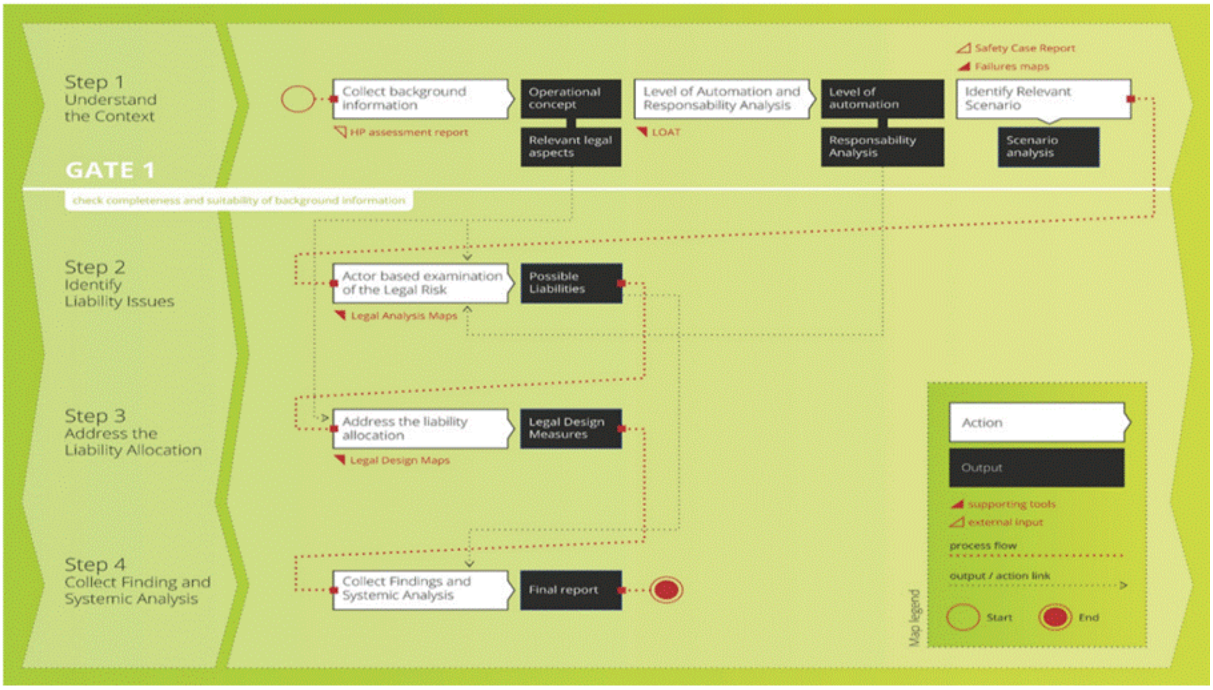

Figure 27: The Legal Case Process

White rectangles represent actions, i.e., sub-steps within each step of the Legal Case. Black rectangles represent a flow of objects from one activity to another, that is, the flow of the information produced in each sub-step of the Legal Case. Bold arrows represent the main workflow. Light arrows represent other connections between objects and actions, that is, the information used as an input for each sub-step. The legal argumentation maps used in the process (Failures maps, and the complete set of Legal Analysis maps) are also inputs and appear as red triangles.

Usually, the Safety Case Report and the Human Performance Assessment (HPA) Report are external inputs and appear as white triangles, meaning that – in case those reports are not available - the Legal Case can be applied without using them. Actually, should the Legal Case be completed before the Safety Case and/or HPA Case, it can also be considered an input for them.

B.2.3 Legal and regulatory framework for taxiing operations

Taxiing, as a key phase of ground operations, involves multiple actors and raises liability issues typically assessed on an individual basis. However, given the complexity and interdependence of tasks, a contextual and integrated analysis of the general regulatory framework is essential.

The following section outlines the main legal sources governing these operations. Worth to remarked that aviation law is primarily rooted in international treaties, which aim to ensure uniform regulatory standards. Signatory states are responsible for transposing these norms into national law and ensuring their consistent application. Within the EU, these international principles are further shaped by the Union's harmonization goals. Through initiatives such as the Single European Sky (SES), EU law refines and adapts international rules to fit the Union's integration strategy. Nevertheless, liability regulation largely remains within the purview of Member States, with EU legislation mainly addressing private law and safety standards[7].

B.2.3.1 Chicago Convention and Annexes

The Chicago Convention (ICAO) of 1944[8] explicates the principle of national sovereignty in international aviation law. According to this convention, national States have "complete and exclusive sovereignty over the air space above its territory" (article 1). As a consequence, according to the following article 28, each national State "undertakes, so far as it may find practicable, to: (a) provide, in its territory, airports, radio services, meteorological services and other air navigation facilities to facilitate international air navigation, in accordance with the standards and practices recommended or established from time to time, pursuant to this Convention [and] (b) adopt and put into operation the appropriate standard systems of communications procedure, codes, markings, signals, lighting and other operational practices and rules which may be recommended or established from time to time, pursuant to this Convention".

In light of this, national States may be liable for the mismanagement of their airspace (and the respective necessary services) even when they have delegated their functions to air navigation service providers. In the absence of an international regime, eventual judicial decisions are regulated according to the law applicable where an accident occurred.

As per the Annex 2 of the ICAO Convention (Rules of the Air)[9], taxiing is intended as any "movement of an aircraft on the surface of an aerodrome under its own power, excluding take-off and landing".

These operations, qualified as surface movements of the aircraft (Annex 2), shall take place along taxiways, namely "a defined path on a land aerodrome established for the taxiing of aircraft and intended to provide a link between one part of the aerodrome and another". According to the recommendations contained in Annex 11 (Air Traffic Control Services, Flights Information Services, and

[7] This section draws on the scoping studies conducted within the HUCAN project – *Holistic Unified Certification Approach for Novel systems based on advanced automation* (SESAR Exploratory Research, GA ID: 101114762) – specifically Deliverable D3.1 *Certification Methods and Automation: Benefits, Issues, and Challenges* and Deliverable D3.2 *Innovative Approaches to Approval and Certification*. Particular reference is made to the sections addressing: 3.3 *Single European Sky*; 3.4 *EASA Basic Regulation*; 3.4.1 *Requirements for Automated Systems*; and 3.4.3 *ATCO Licences and Certificates*. For a more comprehensive treatment of these topics, the reader is referred to the above mentions documents.

[8] ICAO, Convention on International Civil Aviation. Chicago, Illinois, USA (ICAO, Doc.7300/9).

[9] ICAO, Annex 2 – Rules of the Air, Eleventh edition, July 2024.

Alerting Services)[10], "such routes should be direct, simple and where practicable, designed to avoid traffic conflicts".

In light of the above, Annex 2 states the pilot in command (PIC) is the actor that above all the others is responsible for "taking action, including collision avoidance manoeuvres based on resolution advisories provided by ACAS equipment, as will best avert collision". More specifically, the document prescribes the priority rules as follows:

- "An aircraft taxiing on the manoeuvring area of an aerodrome shall give way to aircraft taking off or about to take off".
- "In case of danger of collision between two aircraft taxiing on the movement area of an aerodrome the following shall apply: a) when two aircraft are approaching head on, or approximately so, each shall stop or where practicable alter its course to the right so as to keep well clear; b) when two aircraft are on a converging course, the one which has the other on its right shall give way; c) an aircraft which is being overtaken by another aircraft shall have the right-of-way and the overtaking aircraft shall keep well clear of the other aircraft."
- "An aircraft taxiing on the manoeuvring area shall stop and hold at all runway-holding positions unless otherwise authorized by the aerodrome control tower".
- "An aircraft taxiing on the manoeuvring area shall stop and hold at all lighted stop bars and may proceed further when the lights are switched off".

On the other hand, considering the position of ATSPs, Annex 11 specifies these latter have the responsibility "to issue clearances and information for the purpose of preventing collision between aircraft under its control and of expediting and maintaining an orderly flow of traffic". This is to highlight how ATSPs, and pilots have shared responsibility – and, as a consequence, complementary liability risks – in the management of these operations.

It is noteworthy that, in this scenario, signatory States of the ICAO Convention generally must implement systematic and appropriate ATS safety management programmes to ensure safety in the provision of the ATS at aerodromes. However, considering the European context, national states here have implemented ANSPs as state-run or independent agencies. Therefore, liability issues concerning these providers may eventually involve even State as delegating subject. The liability of States, even in this legal framework, is primarily regulated by national and bilateral agreements or cross-border provisions.

B.2.3.2 Montreal Convention

[10] ICAO, Annex 11 – Air Traffic Services, Fifteenth edition, July 2018.

Modernizing the rules previously introduced by the Warsaw Convention of 1929[11], the Montreal Convention of 1999[12] currently provides the international law liability framework for international carriage by air. The EU recognized and implemented the regime defined by this Convention as per the Reg. EC 889/2002[13] amending the previous Regulation (EC) No 2027/1997 on air carrier liability.

According to the article 17 of the Convention, "The carrier is liable for damage sustained in case of death or bodily injury of a passenger upon condition only that the accident which caused the death or injury took place on board the aircraft or in the course of any of the operations of embarking or disembarking".

In light of this, carriers and companies can be vicariously liable for the actions performed under the responsibility of the pilot in command and the cabin crew. This responsibility, covering all the events that occurred on board the aircraft, includes even the phases of taxiing. The effects of this exposure, however, are limited to civil liability, imposing adequate insurance requirements. On the other hand, in light of the personnel nature of criminal liability, in case of death or injuries, this latter is usually charged to the individuals.

B.2.3.3 Single European Sky

The EU established a common regulatory framework for airspace management on its territory, adopting a dedicated legislative package known as Single European Sky (SES). This initiative aims at ensuring and promoting the maximum possible regularity, security, safety and efficiency of continental air services. To pursue these objectives, the primary goal of the regulatory strategy is to consolidate and enhance the harmonization of national aviation law in all the domains related to the EU competences.

The structure of ATM in Europe is based on a series of regulations adopted by the EU, starting from 2004, that define the SES framework. However, overtime this framework has evolved and now gravitates around the Reg. (EU) 2024/2803 of the European Parliament and of the Council of 23 October 2024 on the implementation of the Single European Sky (recast)[14].

Key innovations include strengthened oversight mechanisms via national supervisory authorities, the introduction of common information services (CIS) to support U-space integration, enhanced

[11] UNTC. Convention for the Unification of certain Rules relating to International Carriage by Air (with Additional Protocol). Warsaw, Poland, 12.10.1929 (UNTC, Reg. no. 3145).

[12] ICAO. Convention for the Unification of certain rules for international carriage by air. Montreal, Quebec, Canada, 28.05.1999 (ICAO, MC99).

[13] Regulation (EC) No 889/2002 of the European Parliament and of the Council of 13 May 2002 amending Council Regulation (EC) No 2027/97 on air carrier liability in the event of accidents (OJ L 140, 30.5.2002, p. 2–5; ELI: http://data.europa.eu/eli/reg/2002/889/oj).

[14] Regulation (EU) 2024/2803 of the European Parliament and of the Council of 23 October 2024 on the implementation of the Single European Sky (recast) (OJ L, 2024/2803, 11.11.2024, ELI: http://data.europa.eu/eli/reg/2024/2803/oj).

EUROPEAN PARTNERSHIP
Co-funded by the European Union

performance and charging schemes managed by the Performance Review Board, and updated provisions for network management and airspace classification. The Regulation also introduces stricter requirements on the certification, accountability, and financial robustness of ATM/ANS providers, and amends Regulation 2018/1139[15] by integrating key performance, safety, and interoperability provisions under a single regulatory framework.

B.2.3.4 EASA Basic Regulation

Regulation (EU) n. 2018/1139 (Basic Regulation, hereinafter: EBR) entered into force in September 2018 with the aim of updating and consolidating existing aviation safety laws and revising the mandate for the European Union Aviation Safety Agency (EASA).

The main objective of the EBR is to establish and maintain a high uniform level of civil aviation safety, to set out the legal basis for the establishment of EASA, to specify EASA's competencies and to establish the scope of common aviation safety requirements. Its scope includes topics such as airworthiness, aircrew licensing, environmental compatibility related to aircraft operations (including third-country operators), ATM/ANS (including air traffic controllers licensing), aerodromes, ground handling and unmanned aircraft.

In particular, the EBR outlines some guiding principles that measures taken under the regulation must comply with. These principles include:

- Reflecting advancements and best practices in aviation, considering global aviation experiences, and scientific and technical progress.
- Relying on the best available evidence and analysis.
- Allowing for immediate response to established causes of accidents, incidents, and security breaches.
- Considering the interdependencies between different aviation safety domains and other technical areas like cybersecurity.
- Establishing performance-based requirements and procedures while allowing flexibility in compliance methods.
- Promoting cooperation and efficient resource utilisation among Union and Member State authorities.
- Utilising non-binding measures, including safety promotion actions, when feasible.
- Accounting for international rights and obligations concerning civil aviation, including those outlined in the Chicago Convention.

[15] Regulation (EU) 2018/1139 of the European Parliament and of the Council of 4 July 2018 on common rules in the field of civil aviation and establishing a European Union Aviation Safety Agency, and amending Regulations (EC) No 2111/2005, (EC) No 1008/2008, (EU) No 996/2010, (EU) No 376/2014 and Directives 2014/30/EU and 2014/53/EU of the European Parliament and of the Council, and repealing Regulations (EC) No 552/2004 and (EC) No 216/2008 of the European Parliament and of the Council and Council Regulation (EEC) No 3922/91 (OJ L 212, 22.8.2018, p. 1–122; ELI: http://data.europa.eu/eli/reg/2018/1139/oj. Current consolidated version: 01.12.2024 – ELI: http://data.europa.eu/eli/reg/2018/1139/2024-12-01).

EUROPEAN PARTNERSHIP
Co-funded by the European Union

Moreover, measures taken under this Regulation must be tailored to the nature and risk level of specific activities, considering factors such as the presence of non-flight crew individuals onboard, potential risks to third parties or property on the ground, aircraft complexity and performance, flight purpose, airspace usage, operation scale and complexity, the ability of affected individuals to assess and control risks, and past certification and oversight outcomes.

A core part of the EBR provides comprehensive substantive requirements applicable in different areas (airworthiness and environmental protection; aircrew; air operations; aerodromes; ATM/ANS; air traffic controls; unmanned aircraft; aircraft used by a third-counter operator into, within or out of the EU). Depending on the specific area, requirements may refer to:

- Design or performance of the overall product/equipment/infrastructure or any of its part, as well as changes to the design of such product/equipment/infrastructure or their parts (e.g., materials and equipment for airworthiness, noise minimisation for environmental compatibility in airworthiness, aircraft performance);
- Procedures (e.g., air operations or air traffic management);
- Organisational aspects of entities involved in designing, producing, managing, maintaining products or infrastructure, as well as those involved in training personnel (e.g., essential requirements for qualified entities, responsibility for aerodromes management);
- Physical fitness, knowledge and skills of personnel, including licences and training requirements (e.g. pilots, crews, ATCOs).

Accordingly, while the EBR does not directly enter into the details of the liability framework, such legal instrument includes a set of precautionary rules, namely safety rules aimed to prevent possible accidents (and related losses). In principle, precautionary standards and liability rules are closely related to each other. Indeed, liability norms are intended as legal remedies transferring the risk that not prevented beforehand. In particular, during the liability analysis, SES precautionary rules are taken into account to assess whether the conduct of the involved parties was negligent or not compliant with the applicable precautionary measures.

B.2.3.4.1 Requirements for automated systems

Given the scope and objectives of the ASTAIR project, the analysis will primarily focus on the requirements for the solutions intended for ATM in general, and on ground operations in particular. With regard to the requirements applicable to automated systems in the aviation domain, Regulation (EU) 2018/1139 establishes certification obligations related to design activities[16], underpinned by the following key principles:

- Fitness for Purpose: Automated tools that provide information or operational recommendations to users must be designed, developed, and maintained to ensure suitability, reliability, and performance consistent with their intended use in operational environments.
- Communication Integrity: Communication protocols between Air Traffic Services (ATS) units and aircraft, as well as between ATS units themselves, must guarantee that all exchanges are

[16] Especially, Article 12 and 43.

timely, accurate, clear, and unambiguous. These communications must be protected against interference, commonly understood by all relevant stakeholders, and, where applicable, explicitly acknowledged.

- System Safety Architecture: The design of systems and their constituent components, whether considered individually or as an integrated whole, must ensure that the probability of a failure resulting in a total system breakdown is inversely correlated with the severity of its impact on operational safety. This reflects a risk-based, safety-critical systems engineering approach.
- Human Factors Integration: The design of systems and components must systematically account for human performance limitations and capabilities, both in isolation and when operating within integrated environments, thereby supporting effective and safe human-machine interaction.

B.2.3.4.2 Requirements for AI-based solutions

In the context of AI-based solutions, the regulatory principles established under existing aviation safety frameworks must now be complemented by the new provisions introduced by the Artificial Intelligence Act (Regulation (EU) 2024/1689 – hereinafter: AI Act)[17]. Under this Regulation, AI systems that function as safety components of a product, or constitute the product itself, falling within the scope of the EASA Basic Regulation are generally classified as "high-risk" (AI Act, Article 6(1)) For this category, Chapter III of the AI Act prescribes a set of specific compliance obligations. These include: implementation of a risk management system; requirements for data quality and governance; technical documentation and record-keeping obligations; transparency measures and mandatory information to be provided to deployers; human oversight mechanisms; provisions for accuracy, robustness, and cybersecurity; as well as the establishment of a quality management system and post-market monitoring processes.

Notably, Article 108 of the AI Act introduces targeted amendments to the EASA Basic Regulation. It mandates that the European Commission incorporate the essential requirements for high-risk AI systems when adopting implementing or delegated acts in the following domains:

- Airworthiness (Articles 17 and 19 of the Basic Regulation, as amended);
- Certification and oversight of ATM/ANS providers, and of organisations involved in the design, production, or maintenance of ATM/ANS systems and constituents (Articles 43 and 47);
- Regulation of unmanned aircraft systems (Articles 57–58).

In summary, the AI Act is of particular relevance to the aviation sector because:

- It classifies AI systems falling under the EASA Basic Regulation as "high-risk," thereby subjecting them to third-party conformity assessments.

[17] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (OJ L, 2024/1689, 12.7.2024, ELI: http://data.europa.eu/eli/reg/2024/1689/oj).

Co-funded by the European Union

- It does not apply directly to such systems, except through its empowerment of the Commission to adopt implementing and delegated acts aligned with the AI Act's requirements.
- It amends the EASA Basic Regulation to ensure that AI-specific safety, reliability, and compliance criteria are systematically integrated into aviation regulatory processes.

In response to this evolving regulatory landscape, EASA has articulated a set of strategic objectives and anticipated means of compliance aimed at guiding the certification of AI-enabled systems in aviation. Central to this effort is the EASA AI Trustworthiness Framework, which outlines a set of foundational "building blocks" intended to align AI-specific considerations with existing aviation safety and certification practices. These are supported by EASA's concept papers[18] developed under the Artificial Intelligence Roadmap 2.0[19], which articulate strategic objectives and anticipated means of compliance for the certification of AI-enabled systems in aviation.

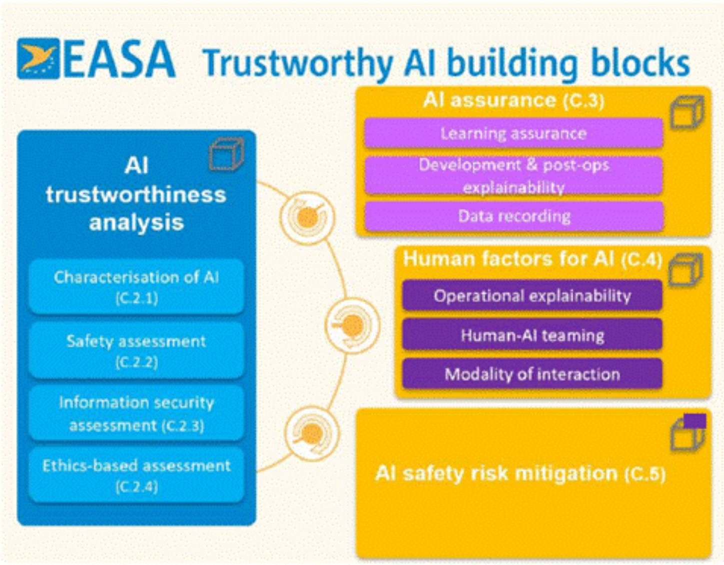

Figure 28: EASA Trustworthy AI building blocks, in EASA (2024). Concept paper. Issue 2, p. 10

B.2.3.4.3 ATCO Licences and Certificates

For personnel licensing, the primary regulatory source is ICAO Annex 1 (Personnel Licensing)[20], where section 4.4 lists the requirements for Air Traffic Controller Licenses. In addition, ICAO published multiple documents related to the duties and procedures for ATCOs, mainly with Doc 9426, "Air Traffic

[18] EASA (2021). EASA Concept Paper: First usable guidance for Level 1 machine learning applications - Issue 01. December 2021. EASA (2024). EASA Concept Paper: guidance for Level 1 & 2 machine learning applications - Issue 02. March 2024.

[19] EASA (2023). Artificial Intelligence Roadmap 2.0 - A human-centric approach to AI in aviation. May 2023.

[20] ICAO. Annex 1 – Personnel Licensing, Twelfth Edition, July 2018.

Co-funded by the European Union

Services Planning Manual"[21], and Doc 4444, "Procedures for Air Navigation Services - Air Traffic Management"[22]. These documents enter the finer details of the operational tasks and methods for ATCOs, like the interaction with the pilot and emergency procedures.

Commission Regulation (EU) 2015/340[23] establishes the technical requirements and administrative procedures for the licensing, certification, and medical assessment of air traffic controllers (ATCOs) and student controllers within the EU. Its core objective is to harmonize ATCO licensing standards across Member States, ensuring high levels of competence, safety, and mobility in civil aviation.

The Regulation introduces uniform rules for the issuance, validation, renewal, suspension, and revocation of ATCO licences, associated ratings and endorsements, and medical certificates. It also sets out certification criteria for training organizations, aero-medical examiners, and medical centres.

Recent amendments[24] have extended the scope to include conversion procedures for military ATCO licences into civil ones, enhanced provisions for inter-State mobility, restructured the rating endorsement system, and introduced requirements related to information security management and incident response (applicable from 2026). These updates aim to modernize the regulatory framework, support workforce flexibility, and strengthen system resilience in line with evolving operational and cybersecurity challenges.

B.2.3.5 Regulatory references for taxiing operations[25]

The European Union has transposed the principles of the ICAO Convention into its legal framework through coordinated action by the EU institutions, the Member States, and EASA. Specifically, EASA has implemented the provisions of the ICAO Annexes by establishing the Standardized European Rules

[21] ICAO. Air Traffic Services Planning Manual (ICAO, Doc-9426).

[22] ICAO. Procedures for Air Navigation Services (PANS) - Air Traffic Management. Sixteenth edition, 2016 (ICAO, Doc 4444).

[23] Commission Regulation (EU) 2015/340 of 20 February 2015 laying down technical requirements and administrative procedures relating to air traffic controllers' licences and certificates pursuant to Regulation (EC) No 216/2008 of the European Parliament and of the Council, amending Commission Implementing Regulation (EU) No 923/2012 and repealing Commission Regulation (EU) No 805/2011 (OJ L 63, 6.3.2015, p. 1–122; ELI: http://data.europa.eu/eli/reg/2015/340/oj. Current consolidated version: 04.08.2024 - ELI: http://data.europa.eu/eli/reg/2015/340/2024-08-04).

[24] Commission Implementing Regulation (EU) 2023/893 of 21 April 2023 amending Regulation (EU) 2015/340 laying down technical requirements and administrative procedures relating to air traffic controllers' licences and certificates (OJ L 118, 4.5.2023, p. 1–65; ELI: http://data.europa.eu/eli/reg_impl/2023/893/oj).

[25] This section draws from and provides an updated perspective on the study carried out within the AEON project – *Advanced Engine Off Navigation* (SESAR Exploratory Research, GA ID: No. 892869) – with specific reference to Deliverable 5.2 *Human Performance Assessment Report* (61).

of the Air (SERA), which form the regulatory basis for safe and harmonized air navigation within the EU.

For the purposes of ASTAIR, considering the context of taxiing operations, the key reference is Commission Implementing Regulation (EU) 2017/373[26], which sets out common requirements for air traffic management (ATM), air navigation service (ANS) providers, and other ATM network functions, including oversight mechanisms. As per the consolidated version, the regulation defines:

- 'Taxiing' as the movement of an aircraft on the surface of an aerodrome or operating site under its own power, excluding take-off and landing (Annex I, point 237);
- 'Taxiway' as a defined path on a land aerodrome intended for taxiing aircraft, providing a connection between different parts of the aerodrome (Annex I, point 238).

Additionally, as per the new regulations (EU) 20/2025[27] on the requirements for the safe provision of ground handling services and for organisations providing them, push-back and towing find specific legal definitions that enable a better understanding also of the liability regime applicable to taxiing operations as envisioned in ASTAIR. Accordingly:

- 'Aircraft pushback' means the movement of an aircraft from a nose-in parking position by using external power of ground support equipment. The operation may involve a towbar (Reg. (EU) 20/2025, Article 3(15)).
- 'Aircraft towing' means the forward movement of an aircraft in service or out of service by using external power of ground support equipment that supports the aircraft's nose landing gear or is attached to it (Reg. (EU) 20/2025, Article 3(14)).

The regulation also provides specific provisions for the movement of vehicles and towed aircraft within the manoeuvring area, to ensure operational safety and air traffic coordination (Annex IV, ATS.TR.240):

- The aerodrome control tower is responsible for controlling the movement of persons or vehicles, including towed aircraft, within the manoeuvring area, as necessary to prevent hazards to landing, taxiing, or departing aircraft (ATS.TR.240(a)).
- During low visibility procedures (LVP), access to the manoeuvring area is to be restricted to essential personnel and vehicles only, with particular emphasis on protecting critical and sensitive areas associated with radio navigation aids (ATS.TR.240(b)(1)).

---

[26] Commission Implementing Regulation (EU) 2017/373 of 1 March 2017 laying down common requirements for providers of air traffic management/air navigation services and other air traffic management network functions and their oversight, repealing Regulation (EC) No 482/2008, Implementing Regulations (EU) No 1034/2011, (EU) No 1035/2011 and (EU) 2016/1377 and amending Regulation (EU) No 677/2011 (OJ L 62, 8.3.2017, p. 1–126; ELI: http://data.europa.eu/eli/reg_impl/2017/373/oj. Current consolidated version: 10.03.2025: ELI: http://data.europa.eu/eli/reg_impl/2017/373/2025-03-10).

[27] Commission Delegated Regulation (EU) 2025/20 of 19 December 2024 supplementing Regulation (EU) 2018/1139 of the European Parliament and of the Council by laying down requirements for the safe provision of ground handling services and for organisations providing them (OJ L, 2025/20, 7.3.2025, ELI: http://data.europa.eu/eli/reg_del/2025/20/oj).

Co-funded by the European Union

- Vehicles operating within the manoeuvring area must adhere to the following hierarchy of right-of-way and instructions (ATS.TR.240(d)):
  - Vehicles and vehicles towing aircraft must give way to aircraft that are landing, taking off, or taxiing;
  - Vehicles must yield to other vehicles towing aircraft;
  - Vehicles must comply with instructions issued by the air traffic services unit;
  - In all cases, vehicles and vehicles towing aircraft must comply with instructions from the aerodrome control tower, regardless of the general right-of-way rules outlined above.

B.2.4 Legal framework for liability

B.2.4.1 Introduction to liability

Before approaching the legal framework concerning liability, it is necessary a terminological premise. Legal scholars and practitioners use to distinguish the consequences of actions or omissions according to different criteria. In this connection, the three keywords in the analysis of the AEON legal framework should be accountability, responsibility, and liability[28].

For the purposes of this report, legal and professional **accountability** within a relational context involves an individual or agency being held to answer for the performance expected by some significant "other". Accountability can furtherly be intended as a principle having a procedural dimension. From an operative perspective, accountability is framed on individual basis, and basically involves: (1) organizational relationship among two or more subjects, defined by law or by factual conditions; (2) a general duty to care about a process or procedure; (3) a general duty to monitor the regular (i.e., correct, and safe) functioning of a process or procedure; (4) a general duty to report and explain the organizational and operative choices related to a process or procedure.

On the other hand, **responsibility** refers to the duty or obligation to carry out a defined task or operation. This duty can be framed on an individual or collective bases, and the subjects involved answer their contribution and its consequences. For the purposes of AEON, responsibility implicitly involves: (1) full personal and situational awareness; (2) adequate professional capacity to carry out the assigned task; (3) relational and contextual understanding of individual contributions and the performance of the procedure taken as a whole.

Finally, **liability** is defined as the condition of being subject to legal consequences deriving from an action or omission. For legal liability to occur, there need to be certain preconditions: (1) a harmful event (2) linked to the action of a person, (3) who was acting in a professional role/task, (4) with no possible justification for the unexpected action. There are also the moral grounds of legal liability that, according to the just culture, should always overlap with legal liability: the person should have moral blame (liability) only when the harm was caused by consciously or recklessly violating a duty/task.

[28] Busnelli, F. D., Comandé, G., Cousy, H., Dobbs, D. B., & Dufwa, B. W. (2005). Basic Norm. In Principles of European Tort Law (pp. 19-22). Springer. https://link.springer.com/chapter/10.1007/3-211-27751-X_2.

Co-funded by the European Union

These three different profiles usually coexist. In some cases, these coincide and are referred to by the same actor. However, in some others, there is no perfect overlap. In these cases, thus, different actors subject to diversified legal regimes may be needed. In particular, those in accountable positions can answer (secondary or vicarious liability) for the action and/or omission of those who took part into the procedures they have to supervise.

B.2.4.2 Types of liability and actors

A single and unique event can raise issues concerning different types of liabilities. In particular, aviation and ATM accidents typically engender:

- **criminal liability**, which presupposes an act (or omission) that violates national criminal legislation and is punished by imprisonment or a fine
- **civil (extra-contractual) liability (or tortious liability)**, based on the intentional or negligent breach of the duty of care, which involves an obligation to redress the loss or injury caused by this breach
- **contractual liability**, which presupposes a breach of contract
- **State/administrative liability** presupposes the violation of a rule or regulation by a public officer who, while exercising their official powers, causes damages or harm
- **product liability** includes the liability of manufacturers and others for defective products
- **organisational liability** is a form of liability of the enterprise for organisational fault in case of injuries caused by commercial activities
- **vicarious liability** refers to the fact that an employer may be held liable for the wrongful act of the employee, performed within the scope of their employment.

The categories of liability presented above can have a different impact on the different categories of operators that may be involved in an accident. Following classes of actors can be distinguished:

- **Physical persons**: the individuals who are directly involved in the provision of air services, namely, pilots, air traffic controllers (ATCOs) or managers of air services
- **Air carriers**
- **ANSPs**
- **Other service providers and actors**: bodies which support the provision of air services, such as technology manufacturers, airport operators, maintenance service providers, certification authorities, national supervisory authorities
- **States**
- **Insurance companies**

Since liability issues usually concern individuals, the report details the liability profile of each subject involved. As shown above, if legal persons basically can incur organisational and vicarious liability, those mainly exposed to criminal liability are the natural persons that materially perform the different tasks.

Aviation, however, experienced peculiar criminal offences. Usually, incidents and casualties are due to accidental situations that the involved operators can difficulty predict or control. Intentional wrongdoings are minimal and quite remote. Instead, the recurrent subjective element in events of this kind (accidents or incidents) usually refers to negligence, recklessness, or malpractice (including inexperience).

EUROPEAN PARTNERSHIP
Co-funded by the European Union

These considerations, therefore, suggest extending the scope of the analysis even to indirect criminal liability issues related to organisational and training gaps and deficits. Inadequate ex-ante and ex-post estimations of each operator's workload, as well as the lack of specific training sessions, may have detrimental consequences on the personal and professional capacities of the involved subject. And these organisational deficiencies can materially influence the state of mind of the actors performing their tasks.

B.2.4.3 AI and liability in aviation: open issues29

As anticipated, the EU has adopted the AI Act (Reg. (EU) 2024/1689) as the cornerstone of its strategy for trustworthy AI. Rooted in a risk-based approach, the regulation imposes proportionate obligations depending on the potential societal and individual risks posed by different AI systems. However, formal compliance does not necessarily equate to immunity from liability. Systems may still cause harm due to malfunction or misuse, despite meeting regulatory requirements[30].

Three core challenges complicate liability attribution in the context of AI:

- **Opacity and Unpredictability**: AI systems, especially those using machine learning and neural networks, often lack full explainability. While aviation systems are required to be predictable, AI inherently introduces functional opacity, making it difficult to trace or justify specific outcomes[31].
- **Multiplicity of Actors**: AI development and deployment involve a distributed ecosystem of stakeholders, designers, programmers, deployers, and operators, each contributing to different phases of the system lifecycle. This complexity renders attribution of responsibility legally ambiguous[32].
- **Hybrid Socio-Technical Systems**: Modern AI systems increasingly function within human-AI teaming environments, designed to augment rather than replace human decision-making.

[29] This section draws on the studies conducted within the HAIKU project – *Human-AI Teaming Knowledge and Understanding for Aviation* (Horizon Europe, GA ID: 101075332) – with particular reference to Deliverable D7.1 *State of the Art in Safety, Human Factors, and Security (SHS) Assurance Processes in Aviation* and Deliverable D7.4 *Recommendations for Liability by Design*, to which the reader is referred for a more detailed treatment of the topic.

[30] Buiten, M., de Streel, A., & Peitz, M. (2023). The law and economics of AI liability. Computer Law & Security Review, 48, 1-20. https://doi.org/10.1016/j.clsr.2023.105794

[31] Bertolini, A. (2020). Artificial Intelligence and Civil Liability. Policy Department for Citizens' Rights and Constitutional Affairs European Parliament. Bertolini, A. (2025). Intelligenza artificiale e responsabilità civile. Problema, sistema, funzioni. Il Mulino.

[32] Expert Group on Liability and New Technologies (Ed.). (2019). Liability for artificial intelligence and other emerging digital technologies. Publications Office of the European Union. 10.2838/573689.

Co-funded by the European Union

While improving efficiency, this coupling may blur accountability lines, especially when humans act as fallbacks during failure events[33].

These factors make the assessment of liability risks quite difficult in practice, as the nature of technology use and the distributed involvement of many stakeholders make it difficult to determine responsibility and accountability for AI-driven outcomes[34].

B.2.4.3.1 AI, defectiveness and product liability

From a legal perspective, the primary reference for addressing these complexities is the newly enacted Product Liability Directive (Dir. (EU) 2024/2853 – hereinafter: New PLD)[35], which expands traditional liability frameworks to include, among other products, software and AI systems (New PLD, Article 4(1)). According to this piece of legislation, software or its components can be considered defective if they fail to provide the level of safety that individuals are entitled to expect or that is required by applicable laws and standards (New PLD, Article 7(1)).

In light of this, according to Article 7, "in assessing the defectiveness of a product, all circumstances shall be taken into account, including:

- the presentation and the characteristics of the product, including its labelling, design, technical features, composition and packaging and the instructions for its assembly, installation, use and maintenance;
- reasonably foreseeable use of the product;
- the effect on the product of any ability to continue to learn or acquire new features after it is placed on the market or put into service;
- the reasonably foreseeable effect on the product of other products that can be expected to be used together with the product, including by means of inter-connection;
- the moment in time when the product was placed on the market or put into service or, where the manufacturer retains control over the product after that moment, the moment in time when the product left the control of the manufacturer;
- relevant product safety requirements, including safety-relevant cybersecurity requirements;
- any recall of the product or any other relevant intervention relating to product safety by a competent authority or by an economic operator;
- the specific needs of the group of users for whose use the product is intended
- in the case of a product whose very purpose is to prevent damage, any failure of the product to fulfil that purpose" (New PLD, Article 7(2)).

---

[33] Wilson, J. H., & Daugherty, P. R. (2018, July-August). Collaborative intelligence: Humans and AI are joining forces. Harvard Business Review. https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces. Bertolini, 2025.

[34] Expert Group on Liability and New Technologies, 2019, 33.

[35] Directive (EU) 2024/2853 of the European Parliament and of the Council of 23 October 2024 on liability for defective products and repealing Council Directive 85/374/EEC (OJ L, 2024/2853, 18.11.2024, ELI: http://data.europa.eu/eli/dir/2024/2853/oj).

Importantly, compliance does not shield developers from liability. If a compliant product fails due to inadequate communication of its limitations or capabilities, manufacturers may still be liable. Moreover, fault-based liability remains difficult to assess given the technical complexity and operational opacity of AI systems.

This concept of defectiveness is broad, ensuring the rights of those who may be harmed by a defective product. It becomes clear that the risk of liability may arise after compliance with standards have been tested. Without a thorough examination of the interplay between compliance and liability from the early stages of design, these risks may remain inadequately addressed[36].

Other challenges remain to be clarified. Specifically applying fault liability rules - assessing whether negligence or malpractice in the deployment and use of AI has occurred - faces the complex nature of these systems and the way they are implemented and used in operations[37].

B.2.4.3.2 Human oversight, responsibility and liability

Under the AI Act, any entity using AI must comply with the obligations set forth in the regulation and is responsible for the technology it employs, particularly when operating under its authority in a professional context (AI Act, Article 2(4)). Additionally, in line with the human oversight requirement, a designated natural person must monitor the AI system's operation. This individual must possess specific expertise regarding the system's capabilities and limitations and ensure proper risk management throughout its use. This includes the authority to override AI-generated outcomes or even halt the system's operation if deemed necessary (AI Act, Article 14).

The human operator thus bears significant responsibility in managing the AI system, acting as the final safety net for risk prevention. However, this creates an ambiguous situation. On one hand, AI is designed to support - and in some cases replace - humans in specific tasks and processes, with the expectation that professionals will diligently incorporate these systems as an essential aid to their work. On the other hand, they are also required to exercise strong professional discretion, deciding on a case-by-case basis whether to trust the system's outputs based on real-time observations of its performance.

This dual structure generates ambiguities in liability attribution, as operators may be held accountable both when they over-rely on AI outputs and when they choose to disregard them, particularly if their decisions result in adverse outcomes. In high-pressure or complex operational environments, human-AI interaction is further complicated by cognitive phenomena such as automation bias (i.e., the tendency to over-trust automated systems), algorithm aversion (i.e., reluctance to rely on AI-

[36] (Buiten, 2024, 255)

[37] (Botero Arcila, 2024, 8 and 15)

generated recommendations), and decision fatigue, all of which can compromise the operator's ability to exercise sound judgment[38].

From a liability perspective, this raises critical questions, especially when considering how this allocation of responsibility might be assessed ex post in the event of an incident. The boundary between over-reliance and over-confidence is particularly thin in high-pressure operational environments, where decisions must be made quickly and under significant stress[39].

The human operator may ultimately be held accountable both for trusting the AI, if its output later proves misleading, and for trusting their own expertise, if they disregard the AI's recommendation and the outcome turns out to be flawed. This tension highlights a key challenge in AI-assisted decision-making: the difficulty of balancing reliance on automation with human judgment.

Apparently, liability claims arising from high-risk AI systems should not be subject to a special, new legal framework when the harm results from human assessment followed by human act or omission, provided that the AI system merely supplies information or advice that was considered by the relevant human actor. According to this line of reasoning, in these situations the damage should be always traced back to human decision or action, as long as the output of the AI system does not intervene between the human act and the resulting damage. It assumes that causality is no more complex than in cases where no AI system is involved.

B.2.4.3.3 AI and liability risk in aviation

The current aviation system is highly functional, safe, and secure, with a consolidated liability regime in the event of accidents. Over time, international and national law, along with legal precedents, have progressively helped clarify the boundaries of responsibility and liability for actors operating in the sector. These lessons learned today provide valuable guidance in proactively addressing both legal and operational risks, even ex ante. However, the development and deployment of AI may necessitate the creation of new, tailored rules and standards.

According to the AI Act, systems designed for the aviation sector are generally classified as high-risk, primarily due to the industry's stringent safety requirements (AI Act, Article 7(1-2)). For this reason, one of the most pressing priorities is to establish a clear regulatory framework with stable and specific requirements tailored to aviation needs. This includes developing dedicated risk management strategies to ensure that AI remains trustworthy, even when applied to such critical functions (AI Act, Article 108 and 112).

[38] De-Arteaga, M., Fogliato, R., & Chouldechova, A. (2020, April 23). A Case for Humans-in-the-Loop: Decisions in the Presence of Erroneous Algorithmic Scores. CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3313831.3376638

[39] (De-Arteaga et al., 2020, 2)

As documented in the work accompanying the EASA AI Roadmap 2.0 and the deliverables produced in recent years[40], the goal is to define objectives and requirements to support the development and deployment of different AI technologies. A key underlying theme of these initiatives is the impact of AI integration within organizations, both in terms of management and operations.

In this respect, the work done so far by EASA to guide the definition of the EU aviation regulatory framework for AI is based on a classification scheme that adapts compliance requirements according to three levels of AI support. Human-AI interaction should range from well-defined assistance to humans (Level 1), encompassing human augmentation (1A) and human cognitive assistance in decision and action selection, to more integrated forms of human-AI teaming (Level 2), which further differentiates into human and AI-based system cooperation (2A) and collaboration (2B). Beyond these levels, advanced automation (Level 3) is envisioned, where AI-based systems make decisions and perform actions, either safeguarded (3A) or non-safeguarded (3B) by human oversight.

In this framework, as specified in the guidance available so far, the key factor in task allocation is the concept of authority, namely, the ability to make decisions without requiring approval from other members.

For Level 1, it is now well established that full authority rests with the end users, given their capability to actively monitor the tasks assigned to the AI-based system, cross-check every decision, and intervene in any action implemented by the AI system. Solutions based on Level 2 AI (particularly at L2B) may involve partial delegation of authority for end users. The users indeed can actively monitor the tasks assigned to the AI-based system and intervene in any action it implements. However, some decisions are made and actions executed by the AI system with a degree of independence.

The same applies to solutions operating under the so-called safeguarded regime, where end users have only materially limited control (Level 3 AI, especially L3A). In this framework, human agents are expected to oversee the AI-based system's operations, with the ability to override its authority for selected decisions and actions when necessary to ensure the safety and security of operations, upon receiving an alert.

In all these scenarios, from a legal perspective, it is reasonable to assume that operators' responsibilities - as defined by existing regulations, such as the ICAO rules outlining the tasks and duties of PICs, flight crew, ATCOs, dispatchers and any other defined role related with operations - formally remain with the human operator, regardless of the practical role played by AI. However, from a substantial standpoint, the interaction between the human agent and the system is explicitly based on a delegated authority, at least for some tasks.

Consequently, when considering the liability risks associated with AI in this context, it is conceivable to assume that if damage arises from an incident due to a system defect - whether from design flaws, manufacturing issues, or information-related problems regarding the system's functionalities,

[40] EASA (2021). EASA Concept Paper: First usable guidance for Level 1 machine learning applications - Issue 01. December 2021. EASA (2024). EASA Concept Paper: guidance for Level 1 & 2 machine learning applications - Issue 02. March 2024.

capabilities, and limitations - liability could be attributed to the manufacturers. However, determining liability becomes far more complex in cases where the incident is caused by human error or organizational factors, especially when assessing personal responsibility.

B.3 Liability analysis by actor

Contextualizing these considerations within the ASTAIR ConOps, the following paragraphs present a reasonable outline of each actor involved.

B.3.1 Technology providers

B.3.1.1 Liability regime

Manufacturers, when delivering defective goods, are liable to third parties who suffered damages under the regime of product liability, which involves a form of strict (i.e., no-fault) liability with additional exemptions (in particular, for design failures). They may also be contractually liable to the purchasers of their goods and services (failure to provide them up to standard may involve contractual infringement).

Generally, when discussing defects, the focus is primarily on three categories:

- A **design defect** scenario may arise if the design was unreasonably dangerous or unsafe, considering the available knowledge at the time of use, evidence of alternative designs, the unreasonable danger posed by the designs, the availability of safety features, the general safety of the product for intended or foreseeable use, the inherent risks versus the benefits of the design, and the reasonable expectations of users.
- A **manufacturing defect** scenario can occur in cases where the technology had a manufacturing defect, meaning the product did not correspond to the intended design, or when low-quality materials were used in the manufacturing process.
- A **warning defects** scenario occurs when instructions for the safe use of the product are unclear, incomplete, inconsistent or missing. It also occurs when poor information is provided about unintended uses and their hazards in specific circumstances. It also includes the lack of information that creates adequate awareness about previous accidents and their consequences regarding similar aircrafts or tools. Warning defects encompass poor instructions about foreseeable risks of not following instructions are not made clear and the proper use of a product is not adequately explained. Finally, it covers cases where there is a failure to warn of the inherent dangers of a product.

Organizations in charge of maintenance are usually subject to contractual liability towards the purchasers of their services. They are subject to fault (negligence) liability toward third parties.

B.3.1.2 Liability analysis

Given that AI represents the most innovative element of the ConOps, all the tasks analysed are novel compared to those foreseen in standard procedures. According to the -categories outlined above, the following specific risks have been identified.

EUROPEAN PARTNERSHIP
Co-funded by the European Union

Design defects	
Safety features	the interface and system architecture do not sufficiently support human oversight, compromising the effective operator's capacity for autonomous and accountable decision-making
User expectations	the system fails to perform its intended functions in accordance with the reasonable expectations of the user
Manufacturing defects	
Manufacturing defects	the implemented cybersecurity and robustness measures are not adequately aligned with the system's intended purpose, operational context, or potential threat landscape
Quality materials	the dataset used for training, testing, or validation does not adequately reflect the system's goals or real-world application or is not aligned with the system's purpose and usage context
Warning defects	
Unclear instructions	inadequate communication of required input data quality, computing resources, system capabilities and limitations, vulnerabilities, or maintenance requirements.
Unclear foreseeable risks	insufficient guidance on system usage, interpretation of outputs, detection of abnormal behaviour, and associated operational protocols
Proper use unclear	poor communication regarding proper system usage, interaction methods, and the resulting consequences

Table 18: Product Liability - Defects

In light of the analysis conducted on the ASTAIR concept, as well as the involvement of the AI system in the scenarios outlined by the Use Cases and represented in the sequence diagrams, the liability risk for the technology provider (i.e. the AI provider, as defined by the AI Act) can be outlined as follows.

Scenario(s)	**UC1.a**	**UC1.b**	**UC2**	**UC3**	**UC4**	**UC5**	**UC8**	**Tot.**
New tasks	*4*	*4*	*5*	*4*	*4*	*4*	*3*	*28*
Causal dependencies	*7*	*5*	*4*	*3*	*3*	*2*	*3*	*27*
Analysed situations	*11*	*9*	*9*	*7*	*7*	*6*	*6*	***55***
Potential liability risks	*8*	*8*	*7*	*8*	*7*	*6*	*6*	***50***
Legal issue(s)	**UC1.a**	**UC1.b**	**UC2**	**UC3**	**UC4**	**UC5**	**UC8**	**Tot.**
Safety features	7	7	4	7	7	5	4	**41**
User expectations	5	5	6	5	6	6	6	**39**
Manufacturing defects	5	8	6	6	5	6	6	**42**

Quality materials	3	3	3	4	3	3	2	21
Unclear instructions	7	8	7	8	5	6	6	**47**
Unclear foreseeable risks	7	7	4	7	7	5	5	**42**
Proper use unclear	0	0	0	2	2	2	1	7

Table 19: Product liability analysis. Overview

B.3.1.3 Interlinkages with corporate liability

It should be specified that potential liability issues concerning defective products are, in most of cases, complemented by a potential risk of corporate liability. This risk is associated with the proper implementation of the system, the adequate definition of usage procedures, the revision of current taxiing operation protocols, the retraining of personnel, and, last but not least, the correct maintenance of the systems. These considerations will be specifically addressed in the sections dedicated to ANSPs, Air carriers and APTOs (B.3.5.1).

B.3.2 ATCO

B.3.2.1 Liability regime

ATCOs are generally regarded as operators holding an accountable position. Their civil liability typically derives from the contractual relationship with their employer and is complemented by the professional insurance coverage mandated by law. Criminal liability, on the other hand, is closely tied to their duty of accountability. ATCOs may face criminal charges in cases of intentional misconduct or negligence leading to an accident. In particular, accidents resulting in fatalities may lead to charges of manslaughter (non-intentional homicide).

In light of the above, task-related responsibilities should not be interpreted narrowly based on their nominal designation but rather in relation to the overall procedural context. This implies a broader duty of oversight, extending to the proper execution of tasks by other involved actors (e.g., TFM, tug operators).

A liability hypothesis can be confirmed if the following conditions are jointly satisfied: there is an injury to a legally protected interest; there is careless behaviour of the person at stake; and there is a (causal) relation between the behaviour and the injury. Some exceptions or counter arguments may be advanced, e.g., the fact that the person's behaviour lacked will.

Careless behaviour may consist of a careless action or a careless omission. Individual's behaviour is careless when the person took action, and the action was careless. Carelessness is usually determined by assessing whether the action violates the standard of due care, which is the proper behaviour that a professional operator would have been required to follow in the given situation. Such expectations depend on the tasks assigned to the ATCOs, as well as on international and national laws, public or private standards and regulations, or even customs. Individual's omission will be careless when the person failed to take action; the person had a duty to act; and the person's action would have prevented the injury. The content of the duty to act will depend on the tasks assigned to the dispatcher, as well as on international and national laws, public or private standards and regulations, or even customs. ATCOs' professional standards are rigorously and systematically defined by the Air Traffic Control Procedures Manual. In this context, a possible defence may rely on the argument that the

EUROPEAN PARTNERSHIP
Co-funded by the European Union

specific omission attributed to the ATCO was not required under the provisions of the Manual. Indeed, in various legal cases, particularly in the United States, controllers who have acted in compliance with the obligations set forth in the Manual have been exempted from further liability.

B.3.2.2 Liability analysis

Building on this premise, the ASTAIR analysis examined the risk of negligent actions or omissions by ATCOs, particularly Ground ATCOs, in light of the innovations introduced by the ConOps. The assessment covered newly introduced tasks, revised tasks, and existing tasks. The analysis also evaluated whether potential negligence could result in direct harm (i.e., a clear causal link between the conduct and the damage) or indirect harm (i.e., the ATCO owed a duty of care to the injured party and there was sufficient proximity between the conduct and the harm suffered).

Scenario(s)	UC1.a	UC1.b	UC2	UC3	UC4	UC5	UC8	Tot.
New tasks	5	4	4	7	3	2	1	26
Revised tasks	2	0	1	0	0	3	1	7
Current tasks	1	1	4	1	0	0	1	8
Causal dependencies	5	1	3	1	1	5	1	17
Analysed situations	8	4	7	7	3	9	3	41
Potential liability risks	12	6	12	9	4	10	4	57
Legal issue(s)	**UC1.a**	**UC1.b**	**UC2**	**UC3**	**UC4**	**UC5**	**UC8**	**Tot.**
Careless action	7	3	4	6	3	6	3	32
Careless omission	3	2	4	5	1	5	1	21
Direct harms	0	0	0	0	0	7	2	9
Indirect harms	8	4	7	7	3	9	3	41

Table 20: ATCO liability analysis - Overview

These results demonstrate that the Ground ATCO, the primary focus of this analysis, is particularly exposed to potential liability risks for the following reasons.

Careless action may occur when the ATCO commits an error in task execution due to negligence, potentially compromising the performance or safety of the system and its associated procedures. In the context of ASTAIR, such occurrences are particularly relevant during taxiing operations in which the ATCO is required to review and approve routing plans, verify operational conditions, authorize procedural steps, delegate and supervise AI-driven functions, monitor traffic and system behaviour, coordinate actions among actors, and manage transitions between phases of control.

Careless omission, on the other hand, refers to situations in which the operator fails to perform a required action due to negligence, potentially affecting system performance or safety. Within ASTAIR , such risks may arise where the ATCO (especially GND ATCO) is expected to provide timely notifications to other actors, actively monitor AI behaviour, acknowledge critical updates or system limitations communicated by the AI, validate or override routing decisions, ensure proper coordination during shifts in responsibility, and confirm key procedural transitions.

It is worth highlighting that certain tasks, as currently defined, are inherently exposed to both careless action and careless omission risks. Their structure and operational criticality make them susceptible to errors in execution as well as to failures in performing required actions, each potentially impacting the overall safety and reliability of taxiing operations.

Regarding the causal link, the data align with the accountability position of the ATCO, who appears to be predominantly exposed to the risk of indirect harm. This is especially due to the causal dependencies between the ATCO's actions and those performed by others, both as inputs to and outputs from their own tasks.

B.3.2.3 Interlinkages with corporate liability

Worth to be noted that, when ATCOs cause damages by not performing their tasks with the required skill and care, their employer (the ANSPs) will have a legal obligation to compensate for the damage based on vicarious liability. This is why, as far as civil aviation is concerned, physical persons, in particular ATCOs, are usually held liable to repair the damage only in connection with a criminal conviction; in all other cases, vicarious liability applies. Finally, ATCOs may be subject to disciplinary sanctions towards their employers for violating their professional duties.

Accordingly, potential issues concerning professional liability of ATCOs should be read in combination with the related corporate liability risks. These are associated with the proper implementation of the system, the adequate definition of usage procedures, the revision of current taxiing operation protocols, the retraining of personnel. These considerations will be specifically addressed in the sections dedicated to ANSPs, Air carriers and APTOs (B.3.5.1).

B.3.3 Tug Fleet Manager

B.3.3.1 Liability regime

To delineate the legal and liability framework applicable to the Tug Fleet Manager (TFM), reference is made to the findings of the AEON project (SESAR Exploratory Research – GA ID: 892869). As highlighted in that project outcomes, the TFM represents a recently introduced operational role which, to date, is not yet supported by a clearly defined regulatory framework[41].

Within the AEON ConOps, and based on the specific set of functions and responsibilities attributed to the TFM, two analogous legal classifications were examined. The first considered the TFM in relation to the role of a dispatcher, focusing on coordination and operational support activities. The second explored the possibility of aligning the TFM's role and competencies with those of an ATCO, due to the

[41] AEON (2022). D5.2 - Human Performance Assessment Report, 68-71 (https://www.aeon-project.eu/wp-content/uploads/2022/12/D5.2-Human-Factors-Assessment-Report.pdf).

EUROPEAN PARTNERSHIP
Co-funded by the European Union

operational similarities involved in the management of a semi-autonomous taxiing fleet during ground movement operations.

From a legal perspective, these two classifications lead to different implications regarding regulatory requirements and liability exposure. In general, the conditions for establishing liability remain consistent with those previously outlined: an injury to a legally protected interest, negligent behaviour, and a causal link between the behaviour and the injury. However, the legal characterization of the TFM plays a critical role in defining both the scope of responsibility (or accountability) and the expectations related to professional conduct, particularly concerning the duty of care and the risk of incurring in negligent actions or omissions.

If classified as a type of dispatcher, the TFM should be regarded as non-aircrew personnel with line responsibility for supporting the safe and timely departure of flights. From a liability standpoint, this role reflects a flexible professional profile, with non-specific accountability and task-related responsibilities determined by the individual's assigned duties. On this basis, the TFM would fall under a civil liability regime governed by the terms of their employment contract and covered by the vicarious liability of the employer (presumably the airport operator, aircraft operator, or ground handler, depending on the adopted economic model). On the other hand, from a criminal liability perspective, the TFM in this configuration would be subject to the general standard of negligence and assessed according to a professional duty of care. As such, the TFM would not hold a clearly defined accountability position. It is therefore reasonable to assume that, under this qualification, the TFM is responsible for the execution of their specific tasks, but has only limited proactive duties concerning procedures carried out by actors not directly under their instruction.

Conversely, if the TFM is considered functionally equivalent to an ATCO, the civil liability regime would still be grounded in the contractual relationship between employer and employee but would also require coverage through statutory professional insurance. The criminal liability framework, however, would be significantly influenced by the stronger accountability obligations associated with such a role. Responsibilities would extend beyond nominal task execution and encompass the entirety of relevant procedures, including a general duty to supervise the proper performance of other involved personnel (e.g., tug drivers). As a result, the TFM could assume a sui generis accountability position. Beyond the direct execution of assigned tasks, they may also be held responsible for the actions of other operators, subject to an overarching duty of care and an expectation of proactive risk monitoring and mitigation.

Following the analysis conducted within the AEON project, it was recommended to classify the TFM as an ATCO-equivalent figure. This approach was intended to enable a more effective allocation of responsibilities, particularly in relation to the accountability duties attributed to both ATCOs and pilots. These considerations are also validated in the context of the ASTAIR project.

Although the current level of project maturity does not allow for a comprehensive exploration of the legal and operational status of this role, the quantitative analysis of scenarios involving legal exposure has been carried out under the assumption that the TFM should be held accountable for the operations under their management.

These considerations are, in principle, consistent with the provisions set forth in Delegated Regulation (EU) 20/2025, particularly in relation to the requirements for the safe provision of ground handling services and for the organisations delivering such services. They also align with the framework for

Co-funded by the European Union

cooperative safety oversight of ground handling activities at the European level among national competent authorities, as established under Implementing Regulation (EU) 23/2025.

In this context, particular attention should be given to the regulatory flexibility introduced to support the harmonisation of procedures at the European level, including the potential integration of new or evolving practices, such as those envisioned under the ASTAIR initiative. This includes, among other aspects, the definition of the responsibilities of ground handling organisations (ORGH.GEN.110); the management of changes, to be applied when modifications to existing procedures are required or deemed appropriate (ORGH.GEN.130); the development and maintenance of the ground handling manual (ORGH.DOC.110); and the implementation of a structured training and assessment programme for personnel (ORGH.TRG.100).

B.3.3.2 Liability analysis

Building on this premise, the ASTAIR analysis examined the risk of negligent actions or omissions by TFMs in light of the innovations introduced by the ConOps. The assessment covered newly introduced tasks, revised tasks, and existing tasks. The analysis also evaluated whether potential negligence could result in direct harm (i.e., a clear causal link between the conduct and the damage) or indirect harm (i.e., the pilots owed a duty of care to the injured party and there was sufficient proximity between the conduct and the harm suffered).

Scenario(s)	UC1.a	UC1.b	UC2	UC3	UC4	UC5	UC8	Tot.
New tasks	3	2	0	0	2	1	1	9
Revised tasks	0	0	0	0	0	3	0	3
Current tasks	1	1	0	0	1	0	2	5
Causal dependencies	1	3	0	0	1	3	2	10
Analysed situations	**3**	**4**	**0**	**0**	**3**	**7**	**4**	**21**
Potential liability risks	**5**	**6**	**0**	**0**	**4**	**7**	**5**	**27**
Legal issue(s)	**UC1.a**	**UC1.b**	**UC2**	**UC3**	**UC4**	**UC5**	**UC8**	**Tot.**
Careless action	3	4	0	0	3	4	2	16
Careless omission	1	2	0	0	4	4	2	13
Direct harms	0	0	0	0	0	2	1	3
Indirect harms	3	4	0	0	3	7	4	21

Table 21: TFM liability analysis. Overview

As the data demonstrate, within the current ConOps configuration and the sequence diagrams of the UCs, the TFM is exposed to a moderate level of liability risk, both in cases of active and omissive careless.

Careless action may occur when the TFM commits an error in task execution due to negligence, potentially compromising the performance or safety of the system and its associated procedures. In the context of ASTAIR, such occurrences are particularly relevant when the TFM is required to reallocate and direct the TaxiBot driver toward the (un)load area for a priority aircraft, re-organise the tug fleet schedule in light of operational priorities, verify safety conditions to delegate authority to the AI system, or acknowledge degraded service levels or updated routing strategies provided by the AI.

Careless omission, on the other hand, refers to situations in which the TFM omits or fails to perform a required action due to negligence, potentially affecting system performance or safety. Within ASTAIR, such risks may arise where the TFM is expected to acknowledge changes in aircraft priority defined by the GND ATCO, analyse warnings issued by the AI, or assume manual control of routing clearances from the AI.

Regarding the causal link, the data align with the TFM's (assumed) accountability position, indicating that the TFM is predominantly exposed to the risk of indirect harm.

B.3.3.3 Interlinkages with corporate liability

Potential issues concerning professional liability of pilots should be read in combination with the related corporate liability risks. These are associated with the proper implementation of the system, the adequate definition of usage procedures, the revision of current taxiing operation protocols, the retraining of personnel. These considerations will be specifically addressed in the sections dedicated to ANSPs, Air carriers and APTOs (B.3.5.1).

B.3.4 Pilots

B.3.4.1 Liability regime

Pilots are generally regarded as operators holding an accountable position. Their civil liability typically derives from the contractual relationship with their employer and is complemented by the professional insurance coverage mandated by law. Criminal liability, on the other hand, is closely tied to their duty of accountability. PICs may face criminal charges in cases of intentional misconduct or negligence leading to an accident. In particular, accidents resulting in fatalities may lead to charges of manslaughter (non-intentional homicide).

Similarly to ATCOs, therefore, their task-related responsibilities should not be interpreted narrowly based on their nominal designation but rather in relation to the overall procedural context. This implies a broader duty of oversight, extending to the proper execution of tasks by other involved actors (e.g., push-back or TaxiBot operators).

Generally, a liability hypothesis can be confirmed if the conditions illustrated above are satisfied (i.e., an injury to a legally protected interest; careless behaviour; and a causal relation between the behaviour and the injury). A careless behaviour can have an active (action) or passive (omission) nature. However, when the analysis converges on pilots there are specific professional references, dedicated to their professional outline. Professional expectations depend on the tasks assigned to the pilot, as well as on international and national law (such as codes of navigation) public and private standards and regulations, or even customs and caselaw. The contents of the duty to act will depend on the tasks assigned to the pilot, as well as on international and national law (such as codes of navigation) public and private standards and regulations, or even customs.

B.3.4.2 Liability analysis

Co-funded by the European Union

Building on this premise, the ASTAIR analysis examined the risk of negligent actions or omissions by pilots in light of the innovations introduced by the ConOps. The assessment covered newly introduced tasks, revised tasks, and existing tasks. The analysis also evaluated whether potential negligence could result in direct harm (i.e., a clear causal link between the conduct and the damage) or indirect harm (i.e., the pilots owed a duty of care to the injured party and there was sufficient proximity between the conduct and the harm suffered).

Scenario(s)	UC1.a	UC1.b	UC2	UC3	UC4	UC5	UC8	Tot.
New tasks	3	2	2	2	0	0	0	9
Revised tasks	0	0	0	0	0	1	0	1
Current tasks	3	0	6	1	0	1	3	14
Causal dependencies	2	2	3	0	0	3	4	14
Analysed situations	**4**	**3**	**3**	**2**	**0**	**4**	**0**	**16**
Potential liability risks	**8**	**4**	**11**	**3**	**0**	**5**	**0**	**31**
Legal issue(s)	**UC1.a**	**UC1.b**	**UC2**	**UC3**	**UC4**	**UC5**	**UC8**	**Tot.**
Careless action	4	1	2	1	0	0	0	8
Careless omission	1	2	2	2	0	4	0	11
Direct harms	4	3	3	2	0	2	0	14
Indirect harms	4	1	3	2	0	3	0	13

Table 22: Pilot liability analysis. Overview

The data indicate that the Pilots are exposed to a marginal level of potential liability risk. Fewer than half of the analysed scenarios are potentially critical, and of these, only a subset is exposed to the risk of careless behaviour, with the following specificities.

Careless action may occur when the pilot commits an error in task execution due to negligence, potentially compromising the performance or safety of the system and its associated procedures. In the context of ASTAIR, such occurrences are particularly relevant when the pilot is required to verify operational conditions, follow the instruction provided by AI-system and supervise its functions, monitor traffic and system behaviour, coordinate actions among actors.

Careless omission, on the other hand, refers to situations in which the operator fails to perform a required action due to negligence, potentially affecting system performance or safety. Within ASTAIR , such risks may arise where the pilot is expected to provide timely notifications to other actors, acknowledge critical situations or system limitations.

It is important to note that certain tasks, as currently defined, are inherently susceptible to both careless action and careless omission risks. The same applies to the causal dimension. The data are consistent with the accountability role assigned to pilots and their frontline position in operations, indicating that this category of actors may be exposed to both direct and indirect harm.

B.3.4.3 Interlinkages with corporate liability

Worth to be noted that, when pilots cause damages by not performing their tasks with the required skill and care, their employer (generally, air carrier) will have a legal obligation to compensate for the damage based on vicarious liability. This is why, as far as civil aviation is concerned, physical persons, pilots included, are usually held liable to repair the damage only in connection with a criminal conviction; in all other cases, vicarious liability applies. Finally, pilots may be subject to disciplinary sanctions towards their employers for violating their professional duties.

Accordingly, potential issues concerning professional liability of pilots should be read in combination with the related corporate liability risks. These are associated with the proper implementation of the system, the adequate definition of usage procedures, the revision of current taxiing operation protocols, the retraining of personnel. These considerations will be specifically addressed in the sections dedicated to ANSPs, Air carriers and APTOs (B.3.5.1).

B.3.5 Corporate liability: a brief overview

As previously noted, the liability assessment focused primarily on the Concept of Operations (ConOps) and the sequence diagrams of the UCs, with particular attention to the risks affecting the actors directly involved in operational tasks. However, as also highlighted in the dedicated sections for each actor, liability related to defective products, particularly in the case of potential AI system failures, as well as the professional liability of the various human operators, must also be assessed in light of broader corporate liability considerations. This includes both liability for innovation and vicarious liability for the actions of employees.

In this regard, aviation law provides specific references for the liability regimes applicable to ANSPs and Air Carriers, who are subject to a dedicated legal framework depending on their role and the nature of their operations. However, as also noted by the project's Advisory Board, the APTO and Ground Handling Services Providers (GHSP) plays a significant role in the implementation and management of ASTAIR. The following section provides further clarifications on the liability regime applicable to these entities.

B.3.5.1 Air carriers' liability regime

As anticipated (§ B.2.3.2), the Montreal Convention regulates the liability for airlines in the event of damages to the passengers, luggage, or other goods during international flights. It entered into force in 1999, superseding the previous Warsaw Convention. The Montreal Convention considers two levels of liability in case of injury to the passengers: (1) liability up to 133.100 SDR irrespective of the airline's fault, and (2) liability in excess of such sum if it fails to prove that it (or the servants or agents employed) did not cause the damage for negligence or wrongful acts or omissions (Article 21).

B.3.5.2 ANSPs liability regime

The primary legal reference for the liability regime of air carriers is the Montreal convention. It is worth noting how, over time, no other relevant international or regional liability regime comparable with this latter emerged. Consequently, at present, the liability of ANSPs does not find specific legal references. Additionally, no State has yet implemented any dedicated regulation to cover the liability issues concerning their ANS agencies. Regulation 889/2002/EU (amending Regulation 2027/1997/EU) only adopted the main principles of the Montreal Convention, stating, among the others, that a modern air transport system requires a regime of unlimited liability in case of passengers' decease or bodily injury.

In consideration of the sovereign nature of ANS, most national laws recognise the primary responsibility of the State, even if an independent body provides the services. However, practical modalities are different from one country to another. A second approach places the service provider on the front liability line: in this case, the claims must be brought against the service provider, but the ultimate responsibility of the State remains due to the lack of specific provisions on this matter. In a third approach, when the ANS functions have been delegated to a third party, the State remains liable only for damages caused by its own, direct fault.

As anticipated, the ANSPs are also subject to vicarious civil liability for torts of ATCOs and managers and to enterprise liability for the safe management and supply of their services. Indeed, these agencies shall take due care of the safe arrangement and performance of their activities and procedures. Furtherly, they shall create a safe and efficient work environment minimizing the possibility of accidents. If they are not in with these best organizational and technical standards, the legal persons may answer for primary liability.

B.3.5.3 APTOs and GHSPs liability regime

As anticipated (B.2.3.5), the legal framework for the liability of airport operators (APTOs) in Europe has evolved significantly with the introduction of the first harmonized EU regulations on ground operations. With the adoption of the EASA ground handling regulations (Commission Delegated Regulation (EU) 2025/20 and Implementing Regulation (EU) 2025/23), airport operators now have explicit obligations to ensure the safe coordination and oversight of ground operations at their airports. They must establish effective management systems and cooperate with ground handling service providers (GHSPs) and air operators to maintain high safety standards.

From a civil liability perspective, airport operators may be held liable for damages resulting from negligence or failure to fulfil their oversight duties. Courts assess whether the operator has implemented adequate safety measures and maintained proper infrastructure. In cases where multiple parties are involved in an incident, liability may be shared based on each party’s role and degree of fault.

Criminal liability is generally personal and may apply to individuals in management positions if gross negligence or breach of statutory duties is established. Notable cases, such as the Linate airport disaster, have demonstrated that airport directors and managers can be held criminally responsible when failures in infrastructure maintenance or safety procedures contribute to accidents.

B.3.5.4 Insights on corporate liability risks

These potential liability issues may involve the airport managing entity as a whole, as well as, more specifically, the companies employing the various operators involved, such as air carriers, ANSPs, and ground handling service providers.

Due to the current level of maturity of ASTAIR’s Concept of Operations (ConOps) and the scope of the validation activities, it is not yet possible to conduct a comprehensive legal risk analysis of entities potentially exposed to corporate liability. Nevertheless, some preliminary insights into possible risk scenarios can be derived from the outcomes of the analysis conducted for the main stakeholders.

As mentioned elsewhere in this document, it should be noted that, in some cases, liability profiles related to defective products and the individual professional responsibility of personnel are complemented by a potential risk of corporate liability. This risk is associated with several factors, including proper system implementation, the adequate definition of operational procedures, the revision of existing taxiing protocols, personnel retraining, and, last but not least, the correct maintenance of the systems.

From a systemic perspective, the scenarios primarily considered fall into the broader category of organisational or systemic faults:

- **Deployment issue**: The design of operational procedures or training programmes provided to operators does not adequately support human oversight, thereby limiting their ability to make autonomous and accountable decisions.
- **Deployment issue**: The use of inadequate input data, insufficient computing resources, or improper maintenance practices may compromise system performance and reliability.

B.4 References and selected bibliography

AEON (2022). D5.2 *Human Performance Assessment Report*.

Bertolini, A. (2020). Artificial Intelligence and Civil Liability. Policy Department for Citizens' Rights and Constitutional Affairs European Parliament.

Bertolini, A. (2025). Intelligenza artificiale e responsabilità civile. Problema, sistema, funzioni. Il Mulino.

Buiten, M. C. (2024). Product liability for defective AI. European Journal of Law and Economics, 57, 239–273. https://doi.org/10.1007/s10657-024-09794-z

Buiten, M., de Streel, A., & Peitz, M. (2023). The law and economics of AI liability. Computer Law & Security Review, 48, 1-20. https://doi.org/10.1016/j.clsr.2023.105794

Busnelli, F. D., Comandé, G., Cousy, H., Dobbs, D. B., & Dufwa, B. W. (2005). Basic Norm. In Principles of European Tort Law (pp. 19-22). Springer. https://link.springer.com/chapter/10.1007/3-211-27751-X_2.

Commission Delegated Regulation (EU) 2025/20 of 19 December 2024 supplementing Regulation (EU) 2018/1139 of the European Parliament and of the Council by laying down requirements for the safe provision of ground handling services and for organisations providing them (OJ L, 2025/20, 7.3.2025, ELI: http://data.europa.eu/eli/reg_del/2025/20/oj).

Commission Implementing Regulation (EU) 2017/373 of 1 March 2017 laying down common requirements for providers of air traffic management/air navigation services and other air traffic management network functions and their oversight, repealing Regulation (EC) No 482/2008, Implementing Regulations (EU) No 1034/2011, (EU) No 1035/2011 and (EU) 2016/1377 and amending Regulation (EU) No 677/2011 (OJ L 62, 8.3.2017, p. 1–126; ELI: http://data.europa.eu/eli/reg_impl/2017/373/oj. Current consolidated version: 10.03.2025: ELI: http://data.europa.eu/eli/reg_impl/2017/373/2025-03-10).

Commission Implementing Regulation (EU) 2023/893 of 21 April 2023 amending Regulation (EU) 2015/340 laying down technical requirements and administrative procedures relating to air traffic controllers' licences and certificates (OJ L 118, 4.5.2023, p. 1–65; ELI: http://data.europa.eu/eli/reg_impl/2023/893/oj).

Commission Regulation (EU) 2015/340 of 20 February 2015 laying down technical requirements and administrative procedures relating to air traffic controllers' licences and certificates pursuant to Regulation (EC) No 216/2008 of the European Parliament and of the Council, amending Commission Implementing Regulation (EU) No 923/2012 and repealing Commission Regulation (EU) No 805/2011 (OJ L 63, 6.3.2015, p. 1–122; ELI: http://data.europa.eu/eli/reg/2015/340/oj. Current consolidated version: 04.08.2024 - ELI: http://data.europa.eu/eli/reg/2015/340/2024-08-04).

Contissa, G. (2017). Automation and liability. An analysis in the context of socio-technical systems. i-Lex, 18-45.

Contissa, G., Sartor, G., Laukyte, M., Schebesta, H., Lanzi, P., Marti, P., & Tomasello, P. (2013). Classification and Argumentation Maps as support tools for liability assessment in ATM. SESAR Innovation Days. https://www.sesarju.eu/sites/default/files/documents/sid/2013/SID-2013-41.pdf

De-Arteaga, M., Fogliato, R., & Chouldechova, A. (2020, April 23). A Case for Humans-in-the-Loop: Decisions in the Presence of Erroneous Algorithmic Scores. CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3313831.3376638

Directive (EU) 2024/2853 of the European Parliament and of the Council of 23 October 2024 on liability for defective products and repealing Council Directive 85/374/EEC (OJ L, 2024/2853, 18.11.2024, ELI: http://data.europa.eu/eli/dir/2024/2853/oj).

EASA (2021). EASA Concept Paper: First usable guidance for Level 1 machine learning applications - Issue 01. December 2021.

EASA (2023). Artificial Intelligence Roadmap 2.0 - A human-centric approach to AI in aviation. May 2023.

EASA (2024). EASA Concept Paper: guidance for Level 1 & 2 machine learning applications - Issue 02. March 2024.

Expert Group on Liability and New Technologies (Ed.). (2019). Liability for artificial intelligence and other emerging digital technologies. Publications Office of the European Union. 10.2838/573689.

Green, B. (2022). The flaws of policies requiring human oversight of government algorithms. Computer Law & Security Review, 45, 1-22. https://doi.org/10.1016/j.clsr.2022.105681

HAIKU (2024). D7.1 *State of the Art in Safety, Human Factors, and Security (SHS) Assurance Processes in Aviation* and Deliverable

HAIKU (2025). D7.4 *Recommendations for Liability by Design*

HUCAN (2024). D3.1 *Certification Methods and Automation: Benefits, Issues, and Challenges*

HUCAN (2024). D3.2 *Innovative Approaches to Approval and Certification*

ICAO, Annex 11 – Air Traffic Services, Fifteenth edition, July 2018.

ICAO, Annex 2 – Rules of the Air, Eleventh edition, July 2024.

ICAO, Convention on International Civil Aviation. Chicago, Illinois, USA (ICAO, Doc.7300/9).

ICAO. Air Traffic Services Planning Manual (ICAO, Doc-9426).

ICAO. Annex 1 – Personnel Licensing, Twelfth Edition, July 2018.

ICAO. Convention for the Unification of certain rules for international carriage by air. Montreal, Quebec, Canada, 28.05.1999 (ICAO, MC99).

ICAO. Procedures for Air Navigation Services (PANS) - Air Traffic Management. Sixteenth edition, 2016 (ICAO, Doc 4444).

Kabashkin, I., Misnevs, B., & Zervina, O. (2023, October 25). Artificial Intelligence in Aviation: New Professionals for New Technologies. Applied Sciences, 13(21), 11660. https://doi.org/10.3390/app132111660

Kirwan, B. (2024, April 9). The Impact of Artificial Intelligence on Future AviationSafety Culture. Future Transportation, (4), 349–379. https://doi.org/10.3390/futuretransp4020018

Marti, P., Lanzi, P., Bannon, L., Sartor, G., Contissa, G., & Masutti, A. (2011). Liability and automation: issues and challenges for socio-technical systems. SESAR Innovation Days. https://www.sesarju.eu/sites/default/files/documents/sid/2011/SID%202011-ALIAS.pdf

Regulation (EC) No 889/2002 of the European Parliament and of the Council of 13 May 2002 amending Council Regulation (EC) No 2027/97 on air carrier liability in the event of accidents (OJ L 140, 30.5.2002, p. 2–5; ELI: http://data.europa.eu/eli/reg/2002/889/oj).

Regulation (EU) 2018/1139 of the European Parliament and of the Council of 4 July 2018 on common rules in the field of civil aviation and establishing a European Union Aviation Safety Agency, and amending Regulations (EC) No 2111/2005, (EC) No 1008/2008, (EU) No 996/2010, (EU) No 376/2014 and Directives 2014/30/EU and 2014/53/EU of the European Parliament and of the Council, and repealing Regulations (EC) No 552/2004 and (EC) No 216/2008 of the European Parliament and of the Council and Council Regulation (EEC) No 3922/91 (OJ L 212, 22.8.2018, p. 1–122; ELI: http://data.europa.eu/eli/reg/2018/1139/oj. Current consolidated version: 01.12.2024 – ELI: http://data.europa.eu/eli/reg/2018/1139/2024-12-01).

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (OJ L, 2024/1689, 12.7.2024, ELI: http://data.europa.eu/eli/reg/2024/1689/oj).

Regulation (EU) 2024/2803 of the European Parliament and of the Council of 23 October 2024 on the implementation of the Single European Sky (recast) (OJ L, 2024/2803, 11.11.2024, ELI: http://data.europa.eu/eli/reg/2024/2803/oj).

Sartor, G., Contissa, G., Schebesta, H., Laukyte, M., Lanzi, P., Marti, P., & Tomasello, P. (2013). The Legal Case. In Proceedings of the 3rd international conference on application and theory of automation in command and control systems (ATACCS-2013) (pp. 96-105). ACM. 10.1145/2494493.2494506

Save, L., & Feuerberg, B. (2012). Designing Human-Automation Interaction: a new level of Automation Taxonomy. Human Factors: a view from an integrative perspective. Proceedings HFES Europe Chapter Conference Toulouse.

Schebesta, H., Contissa, G., Sartor, G., Masutti, A., Tomasello, P., & Taurino, D. (2015). Design According to Liabilities: ACAS X and the Treatment of ADS-B Position Data (SESAR Innovation Days ed.). https://www.sesarju.eu/sites/default/files/documents/sid/2015/SIDs_2015_paper_39.pdf

UNTC. Convention for the Unification of certain Rules relating to International Carriage by Air (with Additional Protocol). Warsaw, Poland, 12.10.1929 (UNTC, Reg. no. 3145).

Wilson, J. H., & Daugherty, P. R. (2018, July-August). Collaborative intelligence: Humans and AI are joining forces. Harvard Business Review. https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces.

Appendix C RTS Solution Scenario 2 Description

The following guidance material was used during the second validation scenario, during the RTS week. This provided a clear structure to the execution of the scenario, while helped the scenario lead and observers to follow the scenario details and tasks required by the ATCO.

C.1 Setup

Overview of Scenario Setup

- MAS-scenario “EHAM_ASTAIR_high_TA”
- RMO North: aircraft land on RWYs 06 + 36R, depart on RWYs 36L + 36C
- Flight schedule: 17-07-2019, 12:00
- Real-time simulation or scripted fast-time simulation
- All inbound A/C use MET, outbound A/C either MET or TET
- 12 tugs are available, task (re-)assignment done automatically
- Tugs are decoupled either at P6/P7 or J-platform
- Scenario is split into 4 parts, with pauses and discussions in between
- Parameters:
- $\Delta t_{HMI \rightarrow MAS} = 33 \text{ sec}$
- $w_{plng} = 10 \text{ min}$
- $h_{plng} \leq 5 \text{ min}$
- Include CTOT: False
- Include TA: True
- # of tugs: 12
- ACO iterations: 400
- ACO ants: 40

Overview of Covered Use Cases

- 1st part: regular arrival / departure + inbound holding (UC1 + UC3)
- 2nd part: departure with emergency (adapted UC7 + adapted UC8)
- 3rd part: multiple delayed departures, inbound flight must hold remotely (UC2 + UC3)
- 4th part: delayed departure + departing flight with emergency (UC2 + UC8)

Hypotheses

- H1: Controllers rate the level of traffic as high.
- H2: Controllers perceive the traffic as flowing well (i.e. without stop-and-go) and in line with the standard procedures.
- H3: Controllers rate the MAS as capable of handling the traffic in nominal situations (standard arrivals / departures).
- H4: Controllers rate the MAS as capable of handling the traffic in non-nominal situations (e.g. emergency returns).
- H5: Controllers agree with the MAS-generated rerouting in case of disruptions (e.g. emergencies, delayed aircraft).
- H6: Controllers perceive the traffic as safe, i.e. do not see the need to manually intervene.

EUROPEAN PARTNERSHIP
Co-funded by the European Union

- H7: The MAS tools (path visualization + future outlook) improve the situational awareness of controllers.
- H8: Overall, the controllers rate the automation to be helpful in managing the traffic.

C.2 Script

Start of FTS/RTS at 12:00h (1st part, UC1 + UC3)

Task: Observe departures with and without tugs, standard + remote-holding arrivals

Traffic:

- 12 TBs are active, task assignment is automatically done
- ARR: KLM604 (AC-1) landing on RWY 06 at 12:01:00. Its stand E22 is still occupied, so it is sent to the remote-holding platform automatically.
- Once the departing flight ACA825 (AC-37) leaves the stand, KLM604 can taxi to E22.

BREAK at 12:11h. Questions to ATCOs:

- How did you perceive the traffic level?
- Which part of the traffic did you pay attention to? What did you observe?
- How satisfied are you with how the automation handled the traffic?
- Would you have handled the traffic differently? If so, how?
- Would you like to input any information to influence the traffic? Why?
- Should the automation provide you any additional information (e.g. KPIs / times / etc)?

CONTINUE (2nd part, UC8)

At 12:12h, ATCO receives message (AC-37): “ACA825, we have technical issues and need to return to the stand. According to our technicians, it will take around 15mins to resolve the issue.” [reaction has to take place within the next 3mins]

Task: handle the disruption.

BREAK at 12:15h. Questions:

- What was your main concern when this event occurred?
- How did you perceive the situation?
- How did you resolve the disruption?
- Why did you choose to send ACA825 to [stand vs. remote holding]? Which information did you base your decision on?
- Did you miss any information?
- Would you have liked the system to provide any decision-support? If so, which and how?

For continuing the scenario, ACA825 is sent to the remote-holding platform, given low priority over other traffic. The delay before it can depart again is set to 15mins.

CONTINUE (3rd part, UC2 + UC3)

ATCO receives the following messages from ground handling:

- 12:15:00 (AC-34), “ELY338 is delayed by around 15mins”

- 12:16:30 (AC-51), "KLM1333 is delayed by around 20mins"
- 12:18:00 (AC-67), "KLM93U is delayed by around 5mins"

Task: issue the delays.

BREAK at 12:20h. Discuss e.g.:

- How convenient was it for you to issue delays?
- Should the MAS have provided any other information or decision-support?
- How well do you think the MAS handled the overall traffic flow in this situation?
- How satisfied are you with how the automation handled the TOBT delays?
- Would you have handled the delays differently? If so, how?

CONTINUE (4th part, UC2 + UC8)

ATCO receives the message at 12:23:00 (AC-75), "KLM611 is delayed by around 10mins"

ATCO receives the message at 12:40:00 (AC-90), "KLM743, we have technical issues and need to return to the stand. We contacted our technicians already – they believe it will take around 15mins to resolve the issue."

Task: handle the disruptions.

BREAK at 12:50h. Discuss e.g.:

- Can you walk me through how you interpreted the situations?
- Was there anything about the MAS's behaviour that felt unclear or surprising?
- Would you have done anything differently without MAS support?

Scenario is finished. Discuss e.g.:

- What is your overall impression?
- Throughout this scenario, did you feel overloaded, underloaded, or well-balanced?
- How did you perceive the balance of your and the automation roles?
- How did you perceive the interactions with the automation?
- How did the interactions affect your decision-making?
- How realistic were the situations?
- How did you perceive the flow of traffic?
- How safe were the operations managed? Did any safety-related issues/concerns occur?
- Which tools on the HMI did you use, and why? How did you benefit from them? Which functionality / aspect did you miss?

EUROPEAN PARTNERSHIP
Co-funded by the European Union