EXPLORATORY RESEARCH

Final Project Results Report

Deliverable ID:	D7.4
Dissemination Level:	PU
Project Acronym:	MAHALO
Grant:	892970
Call:	H2020-SESAR-2019-2
Topic:	SESAR-ER4-01-2019
Consortium Coordinator:	DBL
Edition date:	02 November 2022
Edition:	00.03.00
Template Edition:	02.00.05

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Authoring & Approval

Authors of the document

Name / Beneficiary	Position / Title	Date
Matteo Cocchioni	Project Member	23/04/2022
Nicola Cavagnetto	Project Member	13/06/2022
Brian Hilburn	WP2 Leader	15/06/2022
Clark Borst	WP4 Leader	15/06/2022
Tiago Nunes	Project Member	17/06/2022
Supathida Boonsong	Project Member	17/06/2022
Carl Westin	WP6 Leader	18/06/2022
Matteo Cocchioni	Project Member	02/11/2022

Reviewers internal to the project

Name / Beneficiary	Position / Title	Date
Brian Hilburn	WP2 Leader/CHPR	15/06/2022
Clark Borst	WP4 Leader/TUD	15/06/2022
Tiago Nunes	Project Member/TUD	17/06/2022
Supathida Boonsong	Project Member/LFV	17/06/2022
Carl Westin	WP6 Leader/LiU	18/06/2022
Matteo Cocchioni	Project Member/DBL	20/06/2022

Reviewers external to the project

Name / Beneficiary	Position / Title	Date

Approved for submission to the S3JU By - Representatives of all beneficiaries involved in the project

Name / Beneficiary	Position / Title	Date
Stefano Bonelli	Project Coordinator	20/06/2022
Stefano Bonelli	Project Coordinator	02/11/2022

Rejected By – Representatives of beneficiaries involved in the project

Name and/or Beneficiary	Position / Title	Date

Co-funded by the European Union

Document History

Edition	Date	Status	Name / Beneficiary	Justification
00.00.01	23/04/2022	Draft	Matteo Cocchioni	Draft created
00.00.02	13/06/2022	Draft advanced	Nicola Cavagnetto	Deliverable improvement
00.00.03	15/06/2022	Draft advanced	Brian Hilburn	Deliverable improvement
00.00.04	15/06/2022	Draft advanced	Clark Borst	Deliverable improvement
00.00.05	17/06/2022	Draft advanced	Tiago Nunes	Internal release
00.00.06	17/06/2022	Draft advanced	Supathida Boonsong	Internal release
00.00.07	18/06/2022	Draft advanced	Carl Westin	Internal release
00.00.08	20/06/2022	Deliverable finalised	Matteo Cocchioni	Final release
00.01.00	20/06/2022	Final Release	Stefano Bonelli	Deliverable approved for submission
00.01.01	11/07/2022	Draft advanced	Matteo Cocchioni	Deliverable reopened for revision
00.01.02	03/10/2022	Draft advanced	Carl Westin	Deliverable improvement
00.01.03	14/11/2022	Draft advanced	Clark Borst	Deliverable improvement
00.02.00	20/10/2022	Final Release	Stefano Bonelli	Deliverable approved for submission
00.02.01	02/11/2022	Draft advanced	Matteo Cocchioni	Deliverable reopened for revision
00.03.00	02/11/2022	Final Release	Stefano Bonelli	Deliverable approved for submission

MAHALO

MODERN ATM VIA HUMAN / AUTOMATION LEARNING OPTIMISATION

This document is part of a project that has received funding from the SESAR3 Joint Undertaking under grant agreement No 892970 under European Union's Horizon 2020 research and innovation programme.

Abstract

This document is the Final Project Result, deliverable D7.4 of the MAHALO project. D7.4 captures administrative and technical activities performed during the entire project, including an assessment of the project achievements toward its R&D goals. It takes in input the outputs of Task 7.5, which main objectives are summarising the project activities including all theoretical and empirical research. Thus, it covers the development and integration of Machine Learning models and psychological data stream, along with the Ecological User Interface.

The document provides an overview of the project, from the technical context to the work performed and the main results, the links to SESAR Programme and the main conclusion and lesson learned. Those last elements are a synthesis of what is described at a deeper level in MAHALO D6.2 Field simulation report.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Table of Contents

1 Executive Summary 7

2 Project Overview 9

2.1 Operational/Technical Context 9

2.2 Project Scope and Objectives 9

2.2.1 The ATM Context 10

2.2.2 The Human Operator 11

2.2.3 The Machine Learning agent and the User Interface 11

2.3 Work Performed 13

2.4 Key Project Results 16

2.5 Technical Deliverables 18

3 Links to SESAR Programme 22

3.1 Contribution to the ATM Master Plan 22

3.2 Maturity Assessment 23

4 Conclusion and Lessons Learned 40

4.1 Conclusions 40

4.2 Lessons Learned 43

4.3 Plan for next R&D phase (Next steps) 44

4.4 Guidelines for future AI systems in ATC 45

5 References 49

5.1 Project Deliverables 49

5.2 Project Publications 49

5.2.1 Further Project Publications 50

Appendix A 51

A.1 Acronyms and Terminology 51

List of Figures

Figure 1. MAHALO ConOps Schematic 10

Figure 2. Human Machine Interaction 13

List of Tables

Table 1. Project Deliverables 18

Table 2. ER Fund / AO Research Maturity Assessment for MAHALO Solution 1 24

Table 3. ER Fund / AO Research Maturity Assessment for MAHALO Solution 2 30

Table 4. ER Fund / AO Research Maturity Assessment for MAHALO Solution 3 35

Table 5. Guidelines for future AI systems in ATC: ML/AI Design 45

Table 6. Guidelines for future AI systems in ATC: Personalisation 45

Table 7. Guidelines for future AI systems in ATC: Transparency 47

Table 8. Guidelines for future AI systems in ATC: Human Computer Interaction 48

Table 9. Guidelines for future AI systems in ATC: General 48

Table 10. Acronyms and Terminology 51

1 Executive Summary

The MAHALO project started from simple questions: In the emerging age of Machine Learning, should we be developing automation that is conformal to the human, or should we be developing automation that is transparent to the human? Do we need both? Further, are there trade-offs and interactions between the concepts, in terms of operator trust, acceptance, or performance?

To answer these questions, the MAHALO team has been, first, defining an ATM Concept of Operations and User Interface on which to base this work (see deliverable D2.2 Concept of Operations report, earlier in this series); Second, the team developed an automated conflict detection and resolution (CD&R) capability, realised in a prototype Machine Learning (ML) hybrid system of combined architectures.

The aim of the current work was to use these foundations to address a more specific research questions, as originally laid out by MAHALO, and to experimentally evaluate, using HITL simulations, the relative impact of *conformance* and *transparency* of advanced AI, in terms of e.g. controller trust, acceptance, workload, and human/machine performance. The broad research question to be addressed has been redefined as:

How does the strategic conformance and transparency of a machine learning decision support system for conflict detection and resolution affect air traffic controllers' understanding, trust, acceptance, and workload of its advice and performance in solving conflicts, and how do these factors (conformance and transparency) interact?

In other words, and as written in the Grant Agreement, MAHALO also intended to reach the following four main **Objectives**:

1. Develop an individually-tuned ML system comprised of layered deep learning and reinforcement models.
2. Couple this to an enhanced en-route CD&R prototype display to present machine rationale.
3. Evaluate in real-time simulations the relative impact of ML conformance, transparency, and traffic complexity, on controller understanding, trust, acceptance, workload, and performance.
4. Define a framework to guide design of future AI systems, including guidance on the effects of conformance, transparency, complexity, and non-nominal conditions.

These objectives are completely reached and accomplished, as described in Section 4.1. Here is a summary of the most important results.

Post -experiment analysis of conformance and transparency effects was challenged in the field study by the fact that scenario and simulation both emerged as extraneous variables that required separate 'fine-grained' analyses. That is, preliminary data analysis led the research team to reject a pooled data approach, and instead to treat combination of Simulation and Scenario as a separate sample. This

EUROPEAN PARTNERSHIP

Co-funded by the European Union

obviously had the effect of reducing sample size and statistical power. Nonetheless, some clear effects and trends emerged.

A statistically significant effect of **conformance** was found on agreement, while neither the pooled data nor the fine-grained breakout data present a clear picture about the interaction between conformance and acceptance. A main effect of conformance was on the other hand observed in reducing workload ratings for one Simulation.

In terms of a relationship between **transparency** and advisory agreement, no main effects reached statistical significance, although an interaction between conformance and transparency was suggested by data trends. In terms of workload the trend resulted unclear may be because of some simulation site effects.

These results are discussed at a deeper level in Chapter 4, as well as in D6.2 Field simulation report.

In general, they suggest clear directions for future research. It would be worth focusing on the potential utility of personalisation and of an adaptive system approach, to improve even more the potential an automation with the authority to not only perform a task, but also to shift task responsibility between human and machine.

2 Project Overview[1]

2.1 Operational/Technical Context

The MAHALO project had had two high-level goals: First, to develop and demonstrate a hybrid (using both Supervised- and Reinforcement Learning) ML capability for detecting and resolving en-route air traffic control conflicts; Second, to assess the impact of such a capability in terms of human performance. In particular, MAHALO focused on two constructs thought to underlie human-AI interaction. The first of these is *conformance*, which the project has defined as the apparent strategy match between human and AI systems. The second construct, *transparency*, refers to the degree to which the system makes its internal processes apparent to the operator. The MAHALO project set out to experimentally manipulate these two constructs, and to explore their main and interactive effects on a broad number of human performance measurements, including automation acceptance, agreement with automated advisories, and rated workload.

2.2 Project Scope and Objectives

As said above, MAHALO aims at developing and demonstrating a hybrid ML system for detecting and resolving en-route air traffic control conflicts, as well as at assessing the impact of such a capability in terms of human performance.

Figure 1 presents a high-level schematic of the MAHALO framework. The main components of this framework include:

- a **human operator** (i.e., Air Traffic Controller, ATCO) interacting with the system;
- a **Machine Learning (ML) agent** that can solve conflicts more optimally or learn from the operator to solve traffic conflicts similar to the operator's individual or group-based problem-solving style; and
- a **User Interface (UI)** based on Ecological Interface Design (EID), which provides the operator an insight into the deeper structure of the work domain as well as the inner workings of the ML agent to afford increased transparency.

The three components should always be considered located in the broader external environment, that is the **ATM context**.

[1] The opinions expressed herein reflect the authors' views only. Under no circumstances shall the SESAR Joint Undertaking be responsible for any use that may be made of the information contained herein.

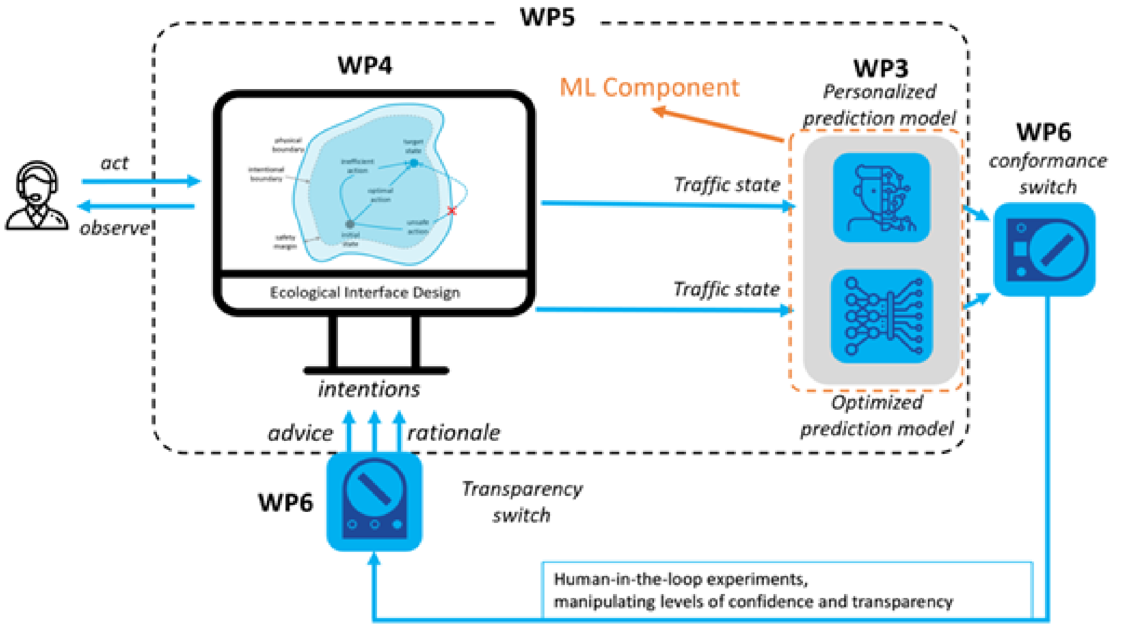

Figure 1. MAHALO ConOps Schematic

2.2.1 The ATM Context

MAHALO targets a future ATM environment consistent with the digital European sky vision 2040 in the European ATM Master plan [SESAR, 2019] and with SESAR and CORUS ConOps for RPAS integration. Central to this vision are the concepts of dynamic airspace organization, flight-centric ATC, and human-centred operations, albeit with increased levels of automation.

Some of the MAHALO project's high-level assumptions regarding the air traffic control task include the following:

- Sectorised airspace, with Tower-, Approach, and ACC regions;
- Data link air-ground communications;
- Sector size and traffic levels to be considerably larger and higher than current (pre-COVID) levels;
- Full ADS-B and SSR Mode S data sharing air-ground (aircraft state, meteo data, etc);
- A future environment consisting of 4DTM, where the majority of separation conflicts have been solved strategically in early planning phases;
- Human-centred (i.e., human-in-the-loop) conflict detection and resolution responsibilities;
- Significant task sharing between human and machine for monitoring and CD&R activities.

Specifying how this task sharing would function, requires examining the role of both the human operator (i.e., the controller) and the machine learning agent.

2.2.2 The Human Operator

The job of an Air Traffic Controller (ATCO) in the ATC conflict detection and resolution (CD&R) task is to ensure proper separation between all aircraft within the sector for which the ATCO is responsible. ATCOs accomplish this by constantly monitoring the traffic situation on a plan view display (e.g. radar or situation display) and by giving commands to aircraft when a potential separation loss is predicted to occur in the future. In addition, ATCOs also strive for efficiency in traffic throughput.

In MAHALO, the human operator (i.e., ATCO) receives assistance from a ML agent to ensure safe separation and increased efficiency in the CD&R task. Because final authority of the CD&R task resides with the human operator, it is essential for successful collaboration between the human and machine agent that the human operator understands the ML system’s advisories and possibly overrides such advisories when deemed necessary.

In this, the human operator and the ML agent will share the same safety goals (i.e., keep aircraft separated at least 5 nm horizontally and 1,000 ft vertically), but may adopt different, and sometimes perhaps conflicting, criteria in achieving those goals. For example, ATCOs adopt human workload management strategies in modulating their control actions, something which is generally irrelevant to computerized agents. There have been many studies conducted on ATCO strategies, which have been reviewed and extensively described in the MAHALO deliverable D2.1.

2.2.3 The Machine Learning agent and the User Interface

The **ML agent** is a main contribution of MAHALO and is envisioned as a teammate to the ATCO. The objective in the design of the ML agent is to increase controller acceptance and cooperation with it. MAHALO focuses on two factors that may foster acceptance: **conformance** and **transparency**. These are illustrated in Figure 1 as two “switches”.

Conformance refers to the ML agent’s ability to solve the CD&R task in ways that align with the individual ATCO’s preferred problem-solving style. To achieve this ability, the automation can “learn” from the operator by being shown data on how the operator solves a given conflict. This approach is useful for two reasons. First, it can be beneficial for the ML agent to learn from individual ATCOs through expert demonstrations (Learning from Demonstration, **LfD**). Second, the system can be taught to use strategies close to what the human operator would usually use. This is **conformal behaviour** and is thought to help foster an operator’s acceptance of automation, in that it is easier for a human controller to understand and accept a solution that is close to what he/she would do himself/herself. **Conformal AI** is, thus, extremely important to the MAHALO project.

Conformal AI is hypothesised to be most useful with novices (as opposed to experts), as a way to foster initial trust and acceptance. Once acceptance and trust is established, a system can propose more optimal solutions that may be less conformal. If the system only proposes the ATCO conformal solutions then it is not really helping the ATCO improve his/her performance and may instead only add to a confirmation bias. Additionally, **Conformal AI** is an important tool but it, in itself, does not provide any further information to the operator about the inner workings of the system. Therefore, it is feasible to consider that the operator might interact with the system for a while and believe he/she knows why it gave a certain suggestion and, yet, no information will be given to verify if the reasoning is correct.

When the ML agents suggest a solution that differs from the individual ATCO’s solution, the ATCO may find it difficult to understand how or why the system suggests this solution. The result may be that the

ATCO rejects the advisory and chooses to disuse the ML agent. To mitigate this problem, **transparency** of the advisory and/or ML agent is important. The goal of transparency is twofold: contribute to both understandability and acceptance of the system. Transparency contributes to understandability by providing the user with more information about the reasoning for a given traffic advisory. Additionally, it also contributes to acceptance since it gives the user a better understanding of the system itself.

MAHALO considered three separate ML models for varying conformance:

- A **Personalized prediction model**, trained on the individual ATCO's CD&R performance that learned and replicated the operator's *subjective* resolution strategies.
- A **Group prediction model**, trained on a group of ATCOs' CD&R performance in a given sample that learned the most dominant (i.e., average) resolution strategies.
- An **Optimized reward prediction model**, that is concerned only with achieving optimality according to the *objective* metrics given to it and does not incorporate ATCO CD&R performance.

The **conformal switch** determines how conformal or optimal the system is at any given time. The **transparency switch** determines what type of information relating to the rationale and intentions underlying an advisory generated by the ML system that will be provided to the ATCO. This allows the system to be versatile and empower the ATCO's with as much information as they might require at any given time. It also allows for experiments to be conducted on exactly how relatively important each of the two features is for performance and acceptance.

The ML concept for the **Personalized prediction model** and **Group prediction model** was based on a concept with Convolutional neural networks used to convert an input, in the form of an SSD or radar display, to a given resolution action by the agent, such as change in heading or speed.

For the **Optimized prediction model,** the team considered several ML approaches, including:

- DQN (Deep Q-Network) -- a variant of the traditional Q-learning algorithm in which a neural network estimates the value of a given state-action pair (Q value) while learning and in later decision making;
- DDPG (Deep Deterministic Policy Gradient) -- combines rule- and reinforcement based models, in such a way that RL selects solutions from the rule base;
- DQfD (Deep Q-learning from Demonstrations)—similar to DQN, DQfD learns a Q function, but it does so use a more efficient *prioritized replay mechanism*. DQfD can learn from human expert demonstrations and achieve better results.

The final decision for the Optimized Model was to develop a Deep Learning agent in such a way that two different approaches can be used: the first approach is DQfD as a proof of concept and the second is DDPG for the main project's experiments.

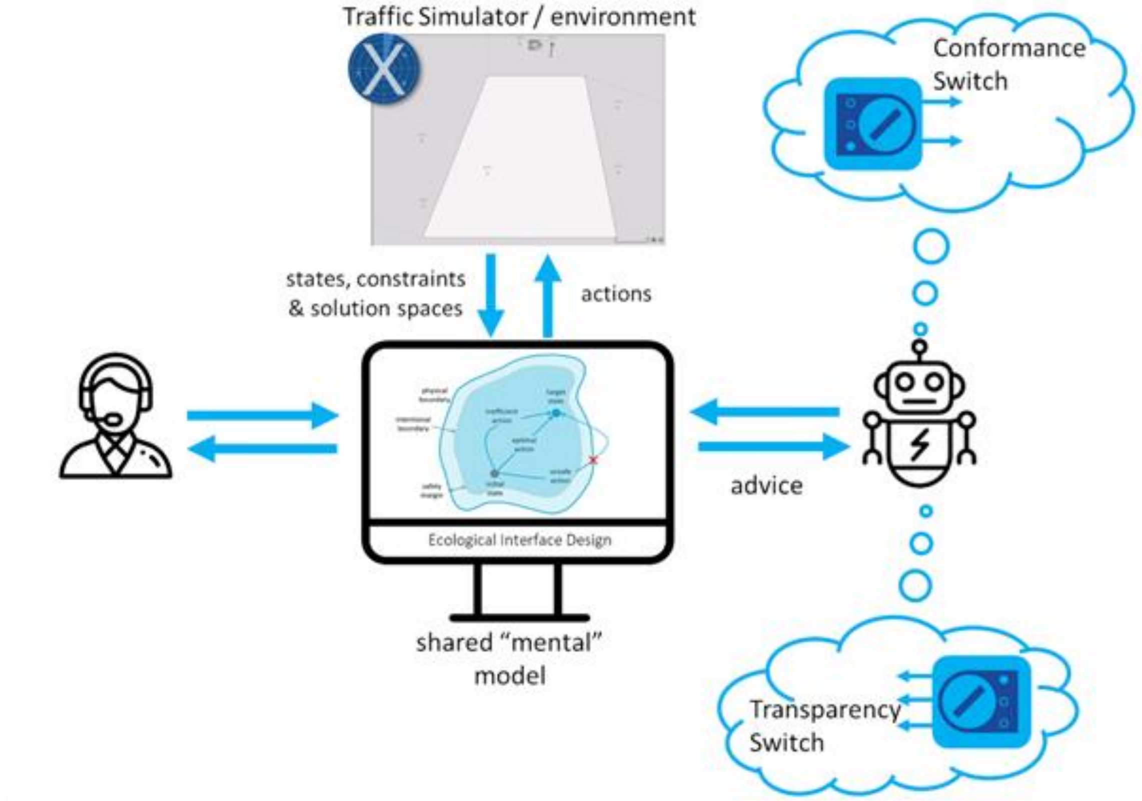

Figure 2. Human Machine Interaction

Figure 2 shows a high-level abstraction of the MAHALO system and ML agent from an ATCO perspective. The ATCO did not know how the ML agent's algorithm is implemented or how it learns. The ATCOs only interaction with the ML agent was through the advice given and the additional explanations that might be given. This means that the ATCO did not require expert training in AI or ML in order to be able to use the system.

2.3 Work Performed

The period covered by the current description is the whole MAHALO project. Here the work performed in this cited period for each technical Work Package.

WP2 focused on three main tasks:

- Conducting a State-of-the-Art Review (SOAR) of the literature on human performance, as it relates to the MAHALO project.
- Conducting a similar review of Machine Learning (ML) concepts and methods (as this is a very dynamic area of research); and
- Formulating an initial definition of the MAHALO Operational Concept - OpsCon).

The first WP2 deliverable was D2.1 Integrated State of the Art Report. D2.1 integrated the T.1 and T2.2 reviews of recent theoretical and empirical research into the areas of human performance and ML, as they relate to the MAHALO project. The integrated D2.1 identified lessons learnt and implications for subsequent MAHALO activities.

The second WP2 deliverable was D2.2 Concept Report. D2.2 was a living document (an updated second edition was submitted in May 2021), that identifies candidate ML approaches, as well as anticipated human roles, information requirements, candidate test protocols, test procedures, experimental design, data collection procedures, and data analysis approaches.

For WP2, all tasks are considered completed and fulfilled. The third edition of D2.2 scheduled for submission in September 2021 was also completed.

In December 2020, WP3 and WP4 were kicked off together, as TUD was the leader of both WPs and the same tasks required joint work. According to the PMP, WP3 and WP4 should have kicked off in September 2020, but the consortium decided to wait for all the inputs coming from D2.2, avoiding developing multiple ML algorithms not backed by a strong theory. There was not a negative impact on the respective activities and deliverables caused by this delay, as it was foreseen.

According to the PMP, the output of WP3 was D3.1 Machine Learning Report and D3.2 Machine Learning Demonstrator, both to be submitted by the end of May 2021 (M12). The priority between the two deliverables was given to the Report to not have an impact on WP5 activities and, although a little bit late, it was submitted before the Intermediate Review Meeting. The Demonstrator was submitted just a few days later, due to its less important impact on other activities.

All major tasks of WP3, particularly, T3.2 and T3.3 were finalised, that is, first versions of the supervised learning model and the reinforcement learning model have been created. T3.1 on integration data streams were also fully finalised; eye tracking hardware was integrated into our system as well as functions to analyse eye behaviour and eye point of gaze data (EPOG) when ATCOs used SectorX during the Simulations. T3.4 validation was also finalised, while the cooperation between SL and RL was done in WP5 alongside their integration into SectorX. T3.5 activities were also completed: the SL model was implemented using a split data set using simulated ATCO data (automation feature of SectorX). The RL model was also implemented using simulated data with different traffic situations. T3.2: For the supervised learning model a CNN approach was chosen, not an LSTM.

In WP4, an initial prototype of the MAHALO Ecological User Interface (E-UI) was developed. The E-UI consisted of several visual representations of traffic conflict states that allow both human air traffic controllers and automated ML agents to perform conflict detection and resolution activities. The visualisations have been inspired by designs developed in previous research activities conducted by TUD, LiU and CHPR. TUD integrated the designs into their Java-based ATC simulator (i.e., SectorX), allowing the tools to be used by controllers as decision-support tools and as ways to monitor automation activities. To validate the designs within the overall MAHALO concept of operations, TUD integrated several rule-based automation algorithms that could perform basic CD&R activities automatically. CHPR and LiU defined the requirements for data recording and validated the usability of the E-UI, the SectorX ATC simulator, and the data recording framework in a small-scale human-in-the-loop experiment. The output of WP4 were D4.1. E-UI design document and demonstrator and D4.2. E-UI validation report. The first one also included a video demonstrator.

WP5 kicked off shortly after the Intermediate Review Meeting: the outcomes from WP3 and WP4 - the ML model and the E-UI - were successfully integrated. WP5 started during month 13 (June 2021) of the project. First, the Consortium integrated the E-UI, outcome for WP4, with the ML model, outcome for WP3. Then, it has been possible to structure the first pre-test to refine scenarios' definition, EPOG capture capability and subjective tests. Timing issues were also addressed. Task 5.3 and Task 5.4 were then finalised, that is, analysing and reporting the activities involved in integrating the ML, E-UI and physiological systems - including the integration decision rationale and the specification of data format, and the analysis and report of Sim 1, including system closed-loop ML training performance, between-participant solution variability, and subjective measures of transparency and conformance.

WP6 started on time during Month 15 (August 2021) of the project. The experimental design was successfully completed and, after a validation Workshop together with the Advisory Board, D6.1 was successfully submitted and approved. During month 17 (October 2021) the definition of the validation activities started. During December 2021 (Month 19), the Conformance pre-test for Cohort 1 of SIM2A was successfully completed both in Padua and Rome, where ATCOs' preferred solutions to solve conflicts using heading were gathered. During January 2022 (Month 20), the Main Experiment of Cohort 1 was successfully accomplished in Padua and Rome. Here, the Consortium gathered data regarding the conformance of AI choices compared with ATCOs' ones, perceived trust, explainability, acceptance and understanding of AI solutions, increased or decreased situational awareness and workload and other important feedbacks related to the use of a similar tool in the workstation. The same pre-test experiment for Cohort 2 took place during March 2022 (Month 22) in Sweden, hosted by LFV. The main experiment for Cohort 2 took place during Month 23 (April 2022) of the project. The CD&R tool in the Swedish trials was also tested with the help of eye-tracking technologies (Tobii2 Glasses owned by LiU) to observe ATCOs exploratory gaze. A final D6.2 Field simulation report has been more recently submitted. The main results and achievements discussed of this report are not included in the current chapter of D7.4 but are summarized below in Chapter 4.

During the project's lifecycle, for WP7, the consortium has participated in some important dissemination events. All of them had been widely promoted using the website and social channels, LinkedIn and Twitter of the project, where a continuous activity is being carried out to promote the results with the main stakeholders. A YouTube account has also been created to collect all the videos of these events and those that will be made in the last moths of the project specifically dedicated to dissemination and exploitation activities. Two video demonstrators from D3.2 and D4.1 have also been included. A project website has been correctly realized and keeps track of all the internal progress of the project and the dissemination activities carried out. Moreover, the D7.5 deliverable has been realized and kept up to date in order to keep track of the most recent dissemination activities performed.

The social channels currently active can be reached at the following links:

- http://mahaloproject.eu/ (the Project website)
- https://twitter.com/H2020_MAHALO (the Twitter Account)
- https://www.linkedin.com/company/h2020-mahalo/?viewAsMember=true (the LinkedIn Account)
- https://www.youtube.com/channel/UCjYW9U2bZxFjLLRVkaUeN9w (the YouTube Channel).

2.4 Key Project Results

Over the project runtime, from WP2 through WP6, a series of deliverables was issued, in step with project technical progress. These deliverables capture the staged process by which the MAHALO team:

- Conducted a state-of-the-art review of Machine Learning (ML) advances (D2.1);
- Developed and demonstrated a Machine Learning (ML) capability (D3.1; 3.2);
- Designed an experimental user interface and simulation capability (D4.1);
- Conducted human-in-the-loop validation trials of the user interface (D4.2);
- Integrated ML capabilities with the simulator and experimental interface (D5.1);
- Conducted a first full simulation to demonstrate the entire test platform (D5.2);
- Specified experimental design for the final simulation sessions (Deliverable 6.1); and
- Conducted a pair of two-phase (Training pre-test, and Main experiment) field studies in which ML models of en-route ATC CD&R were created, and 34 controllers took part in human-in-the-loop trials (D6.2).

The documents describe an extent work performed in the field of research targeting the ATM industry. Here, the increase in the levels of automation is expected to grow in the coming years in all fields of the ATM world, particularly referring to the Conflict Detection and Resolution task. The advent of Artificial Intelligence and specifically Machine Learning algorithms will play a crucial role in this change. So far, what impact these technologies will have on human performance has not been explored, even at a low TRL. The MAHALO project has tried to answer this question, to identify a potential way to increase ATCO's acceptance, through the development of a system able to integrate automatic advisories based on ML and through two field simulations aiming at exploring the effect of machine conformance and transparency on human performance. The MAHALO Project carried out exploratory-level research, achieving a TRL1 for the three solutions for advanced support for Conflict Detection and Resolution by Tactical Controller in en-route: 1. Machine Learning (ML) modelling system supporting the resolution of en-route ATC conflicts; 2. Ecological user interface (E-UI) providing conflict resolution advisory transparency, and 3. Guidelines for the design of future AI systems.

MAHALO conducted two field simulations at two sites between December 2021 and April 2022. In total, 36 participants took part (final n=34 after data from two participants was discarded). Two constructs were hypothesized as critical to the interaction between aircraft controller and ML system: conformance was defined as the similarity between human and machine resolution strategy; transparency was defined as the degree to which the system made clear its underlying rationale. MAHALO conducted field simulations to evaluate the impact of conformance and transparency manipulations on controller acceptance, agreement, workload, and general subjective feedback, among other measures. Each simulation consisted of two phases. First was a training pre-test in which controllers interacted with scripted traffic scenarios that presented two-aircraft closing conflicts, and which recorded controllers' resolution strategies. Second was a main experiment phase, in which the same controllers interacted with ML solved analogues of the pre-test scenarios. ML solutions were developed during an interim training phase, in which several ML models were trained or synthetically generated to output conflict solution advisories.

For the main experiment phase at each site, conformance and transparency were manipulated within participant. Conformance was implemented as either a personal model, a group model, or an optimal model. ML was used to build the group and optimal models, whereas a synthetic approach was used to construct personal models for each participant. Transparency of proposed advisories was defined as either a baseline vector solution display, a prototype Situation Space Diagram (SSD) representation, or a text-based condition that combined SSD with a contextual explanation of the systems rationale (e.g. about target Closest Point of Approach, CPA).

The advisory conformance, or personalisation, of advisories had an impact on controllers' response to advisories, but not in a uniform direction. Although personalized advisories received more favourable responses in many cases, there were also cases when the optimal or group advisories were favoured. There was no strong effect of advisory transparency on controllers' responses. An in-depth analysis was made dividing participants in two groups depending on how close their separation distance preference (i.e. their personal model) was to the target separation distance aimed for by the optimal model's advisory. The analysis revealed a reoccurring pattern emerged where the group of participants, whose average separation distance measured in the training pre-test was closer to the separation distance aimed for by the optimal advisory, showed unchanged or more positive responses to the advisory with increasing transparency. That is, their acceptance of advisories and ratings of agreement, conformance, and understanding was higher compared with the other group.

The project also provided valuable findings and guidelines on how to incorporate conformance and transparent mechanisms of AI solutions to conflict detection and resolution in particular, and to problem solving tasks in safety critical systems in general.

An implicit assumption going into this research was that Transparency fosters understanding, acceptance and agreement. As a thought experiment, however, consider the case where poorly functional automation is outputting advisories. In this case, transparency might have the opposite effect and lower controllers' agreement and acceptance of the system. The notion here is that if transparency involves making clear to the operator the inner workings of the algorithm, it does not necessarily increase agreement and acceptance, but should optimize them. Transparency and explainability should increase acceptance and agreement for an optimal algorithm, which should also decrease acceptance and agreement for a sub optimal algorithm.

Although personalization of ML systems is held as a positive goal, there is one potential challenge that we need to consider. Namely, attempts to personalize advisory systems introduce the risk that they drive the operator to solve the problem in a particular way. For example, the simulated advisories aimed to solve en-route conflicts using a single intervention with only one of two involved aircraft. This approach is inconsistent with controllers who solved the conflict with two interactions (for example, slightly turning both aircraft). It should be noted that the way advisories are framed can give a suggestion for how the system proposed to solve a given conflict, and offers an implicit reference against which controllers' judgment and decision is formed. Without an advisory system the controller would search for information and cues with regard to traffic pattern, speeds, altitude etc. in deciding how to solve a conflict. Past research has noted that advisory systems can have the unintended consequence of increasing task load. The notion is that whereas a current controller has to devise a solution, under an automated advisory system that controller has the additional task of processing the advisory, and comparing that to their own strategy.

Furthermore, during the project's lifecycle, MAHALO participated in some important dissemination events, such as DASC 2020, SIDs 2020, SIDs 2021, ENGAGE Conference September 2021, ICCAS 2022, and two SJU ER4 Automation Workshops, from which it has obtained excellent feedback from the audience. Moreover, MAHALO hosted two public Workshops (in October 2021 and May 2022) with the Advisory Board and relevant ATM stakeholders to validate and gather useful insights regarding the experimental design for Cohort 1 and 2 of the simulations, and to present the first results of both the Italian and Swedish Simulations. After that, the project and its results were also presented at the ANACNA Conference hosted in Rome in April 2022, at the World ATM Congress (with a leaflet) in June

2022 and to EASA in October 2022. For the last months of the project, MAHALO planned to attend EASN 2022 and SIDs 2022 Conferences, for which two scientific papers are already produced or in the process to be produced, at the time of writing.

2.5 Technical Deliverables

Table 1 below lists all the Technical Deliverables realized during the two years of the project, that describe the entire MAHALO process from the initial literature review to the final data analysis and conclusion, after the Simulations were carried out. For each deliverable is provided a brief summary of the contents, with a specific focus on the relationship with the overall project achievements.

A number of confidential, non-technical deliverables about the management of the entire project are not included in the table.

Table 1. Project Deliverables

Reference	Title	Delivery Date[2]	Dissemination Level[3]
Description			
D1.1	Project Management Plan	30/06/2020	CO
The document presented the Project Management Plan (PMP) that complements the project information provided in the Grant Agreement and its Annex I - Description of Action, integrating in particular more detailed procedures, briefly describing the Communication and Dissemination strategy, addressing the Ethics Requirements and implementing any additional refinement agreed at the Kick-off meeting.			
D2.1	Integrated SOAR	01/10/2020	PU
D2.1 Presented an integrated report on the output of human performance and ML reviews, highlighting the latest theoretical and empirical work. The document integrated the outputs of T2.1 and T2.2. Several topics are analysed, from a general overview of the main activities performed by a modern ATCO and the best Conflict Detection and Resolution strategies to the importance of integrating Machine Learning in future normal operations and the pivotal role of MAHALO Project in exploring the real possibilities in realising and applying such a technology. Furthermore, other important questions about human performance were reviewed, in terms of conformance and transparency. The Deliverable can be accessed at the following link: http://mahaloproject.eu/wp-content/uploads/2021/06/MAHALO-D2.1-SOAR-v2.0.pdf			
D2.2	Concept report	30/12/2020	PU
D2.2 reflected the output of T2.3 and built on the earlier D2.1 with additional inputs from WP3 and WP4. It presented the MAHALO ATM concept of operations, including anticipated human roles, information requirements, and candidate test protocols. Test protocols include operational scenarios, test procedures,			

[2] Delivery data of latest edition

[3] Public or Confidential

experimental design, data collection procedures, and data analysis approaches. This report also identified candidate ML architectures to be used in MAHALO.

The Deliverable can be accessed at the following link:

http://mahaloproject.eu/wp-content/uploads/2021/10/MAHALO-D2.2-ConOps-Update-v3.0.pdf

D3.1	ML report	25/06/2021	PU

D3.1 described the main tasks performed in WP3, specifically the development, training and verification of (1) a Supervised Learning (SL) model and agent for providing conformal, or personalised, resolution advisories, (2) a Reinforcement Learning (RL) model and agent, based on rules and constraints, to derive optimal conflict resolution advisories that may be nonconformal to the human ATCO, and a (3) combined SL/RL model and agent that considers human strategies for conflict resolution in deriving an optimal resolution advisory. Moreover, D3.1 described conceptually how the different AI agents should interact and be integrated with the simulator platform SectorX to perform the MAHALO experiments.

The Deliverable can be accessed at the following link:

http://mahaloproject.eu/wp-content/uploads/2022/06/D3.1-Machine-Learning-report_v03.pdf

D3.2	ML demonstrator	06/08/2021	PU

Deliverable D3.2 demonstrated two machine learning models developed in WP3, in the MAHALO project, representing two distinct approaches to automation support. Consequently - supplementing D3.1 – the demonstrator showed our approach to investigate questions on conformance and transparency in terms of one machine learning model based on Supervised Learning to replicate human problem-solving strategies and one model using Reinforcement Learning to provide explanations to the human operator. These models were used in the subsequent human-in-the-loop experiments with air traffic controllers.

The Deliverable and the Demonstrator can be accessed at the following links:

http://mahaloproject.eu/wp-content/uploads/2022/06/D3.2-Machine-Learning-Demonstrator_v02.pdf

https://www.youtube.com/watch?v=pVXTFbrD4T4&ab_channel=MAHALOProject

D4.1	E-UI design doc & demonstrator	31/05/2021	PU

D4.1 detailed the E-UI, which served as a common ground/shared mental model between the human and automated machine learning (ML) agents acting in the same airspace environment. In particular, the interface aimed to add domain and agent transparency to the system, which enabled the human controller to understand what the ML agent is doing and manually intervene if necessary or desired. EID is used as a design framework for achieving the shared mental model. EID emphasises visualising the physical laws and principles governing the ATC work domain, which bounds all actions that can be undertaken by humans and automated agents.

The Deliverable and the Demonstrator can be accessed at the following links:

http://mahaloproject.eu/wp-content/uploads/2021/10/D4.1-E-UI-design-doc-demonstrator_v02.pdf

https://www.youtube.com/watch?v=WJthH_r7LFE&ab_channel=MAHALOProject

D4.2	E-UI validation report	31/05/2021	PU

D4.2 reflected the output of MAHALO T4.3 and described a test of the E-UI SectorX simulator to validate interactions, functions, and output of SectorX were aligned with ML integration and human-in-the-loop simulations requirements. D4.2 also evaluated eye tracking as input tool to the ML system, and as a method for objective performance assessment, with an experienced enroute ATCO. Results of T4.3 confirmed SectorX functionality and simulation realism, validated data logging protocols and formats, and demonstrated the integrated use of eye tracking and output data. Lessons were drawn regarding ATCO strategies, data logging, and eye tracking integration.

The Deliverable can be accessed at the following link:

http://mahaloproject.eu/wp-content/uploads/2021/10/D4.2-E-UI-validation-report_v02.pdf

D5.1	Integration report	13/12/2021	PU

D5.1 described the efforts performed on the integration of Machine Learning models from MAHALO WP3 with the Human Machine Interfaces from MAHALO WP4, as part of MAHALO WP5. A framework was presented that is used to test the integration of the components, containing a number of interfaces where information is transferred between ATC simulators, displays and machine learning models. First the overall pipeline of this integration testing framework was presented. After that the interface and data formats were presented.

The Deliverable can be accessed at the following link:

http://mahaloproject.eu/wp-content/uploads/2022/06/D5.1-Integration-report.pdf

D5.2	Simulation 1 report	13/12/2021	PU

D5.2 described the SIM1 experiment that was carried out in MAHALO WP5: Integration. The purpose of SIM1 was to validate the integrated system, composed of machine learning models and human machine interfaces that was then used for the main MAHALO experiments, carried out in WP6.

The Deliverable can be accessed at the following link:

http://mahaloproject.eu/wp-content/uploads/2022/06/D5.2-Simulation-1-report_v02.pdf

D6.1	Experimental design document	15/10/2021	PU

D6.1 is the Experimental Plan and captured the research team's planned approach to conducting WP5 integration trials, as well as WP6 simulations. D6.1 experimental plan fed WP5 activities (ML and E-UI integration, integration trials of WP5 T5.2) both iteratively and interactively.

This experimental plan followed the structure suggested in SJU's SESAR 2020 Experimental Approach Guidance ER document. The report was the Validation Plan for the broader MAHALO concept and was distinct from the previously submitted D4.2 E-UI Validation Report, which limited itself to evaluation of the SectorX user interface.

The Deliverable can be accessed at the following link:

http://mahaloproject.eu/wp-content/uploads/2022/06/D6.1-Experimental-design-document_v02.pdf

D6.2	Field simulation report	06/06/202	PU

D6.2 represented the field simulation report for both SIM2A and SIM2B performed within WP6 in Italy and Sweden between December 2021 and April 2022 involving 36 participants. Each simulation consisted of two phases. D6.2 details the two phases of each simulation, the conditions used, and discusses results and conclusions.

The Deliverable can be accessed at the following link:

https://drive.google.com/file/d/1IoG056TIDpR567a-Deoo1BfOp6TSg_KX/view?usp=share_link

D7.1	Project Website	02/07/2020	PU

D7.1 developed the Project Website, which can be accessed at the following link.

http://mahaloproject.eu/

D7.2	Workshop 1 Report	16/12/2021	PU

D7.2 contained all the related material to Workshop 1 with the MAHALO Advisory Board. Specifically, Workshop agenda, participant list, workshop presentations and the received feedback from the advisory board. Additionally, it also contains minutes from the workshop and a conclusion & summary of actions section.

The Deliverable can be accessed at the following link:

http://mahaloproject.eu/wp-content/uploads/2022/06/MAHALO-D7.2-1st-Workshop-report_v02.pdf

D7.3	Workshop 2 Report	31/05/2022	PU

D7.3 contained all the related material to Workshop 2 within the MAHALO Advisory Board. In mid-May, the MAHALO project presented the ML models (i.e. SL and RL) and preliminary results of HITL simulations to the

Advisory Board (AB) members and relevant Air Traffic Management (ATM) stakeholders. The ATCOs who have participated in the simulations were also invited to attend the workshop. The Deliverable can be accessed at the following link: https://drive.google.com/file/d/17dlXAoa_jZvBowPEDGvDkSqr-sBaFTOK/view?usp=share_link			
D7.4	Final Project Results report	20/10/2022	PU
The document is the Final Project Result, deliverable D7.4 of the MAHALO project. D7.4 captures administrative and technical activities performed during the entire project, including an assessment of the project achievements toward its R&D goals. It takes in input the outputs of Task 7.5, which main objectives are summarising the project activities including all theoretical and empirical research. Thus, it covers the development and integration of Machine Learning models and psychological data stream, along with the Ecological User Interface.			
D7.5	Communication, Dissemination and Exploitation Plan	16/10/2021	CO
This document described the Communication, Dissemination and Exploitation (CDE) plan and all related activities designed to reach a broad range of stakeholders providing different levels of information and using different communication means, tailored on the basis of the stakeholder role and interest. It also defined the beneficiaries' strategy and concrete actions related to the protection and exploitation of the project results. The deliverable moreover explained how the consortium intends to perform these activities and monitor their impact. The presented activities contribute not only to the external but also to the internal dissemination, among all the partners of the project.			
D8.1	PODP – Requirement No. 4	30/07/2021	CO
This deliverable described the measures taken to ensure that project activities are respectful of human rights, particularly the right to privacy and data protection, and do not generate ethically unwanted personal or social effects. The deliverable described privacy, ethical and other legal concerns, and proposes mitigation measures to address them. The report presented the overview of the regulatory framework concerning the protection of personal data, a short description of the MAHALO research activities involving the processing of personal data, a set of measures that could be implemented to reduce the risk of misuse of the research results.			
D8.2	PODP – Requirement No. 5	31/05/2021	CO
This deliverable described the measures taken to ensure that all the data which had been collected during the Simulation 1, Simulation 2A and Simulation 2B, MAHALO's experiments with students from Delft University, Italian and Swedish ATCOs, had been anonymized and/or pseudonymised.			

3 Links to SESAR Programme

3.1 Contribution to the ATM Master Plan

The MAHALO Project carried out exploratory-level research, achieving a TRL1 for three solutions described in detail in the next section. This can be considered the project's main contribution to the ATM Master Plan. Further in this Chapter a description of the Maturity Assessment for the project Solutions is provided.

In particular, the three Solutions are:

1. Machine Learning (ML) modelling system supporting the resolution of En-route ATC conflicts.

   The solution consists of a Machine Learning (ML) modelling system that explore the effects of ML conformance and transparency, as well as contextual factors (e.g., traffic complexity), on human and en-route ATM system performance.

   The ML system comprised layered deep learning and reinforcement models trained on controllers' performance and control strategies, able to resolve en-route ATC conflicts. It was coupled with an associated Ecological User Interface, which aimed to augment the typical plan view display with machine intent and decision selection rationales. The solution enhances ATCOs' performance, preserving the safety of ATM operations within a given sector when using explainable and trustworthy ML algorithms.

2. Ecological user interface (E-UI) providing conflict resolution advisory transparency

   The solution consists of visual elements in the user interface, based on Ecological interface design, that affords understanding of why a particular conflict resolution solution is recommended. The visual elements increase the transparency of advisories by providing the operator an insight into the deeper structure of the work domain as well as the inner workings of the ML agent.

3. Guidelines for the design of future AI systems

   The solution consists of a series of guidelines for the design of future AI systems, including guidance on the effects of conformance, transparency and complexity. These guidelines have been evaluated through human-in-the-loop simulations considering controller trust, acceptance, workload and human/machine performance.

The three MAHALO Solutions can eventually contribute to achieve Level 1 of Automation, as described in the ATM Master Plan. In fact, in the Simulations the Advisory System always left the human the ability to initiate and execute the action of solving a specific conflict but supported that human in the information acquisition (spotting a conflict), the information analysis (calculating the Closes Point of Approach – CPA) and suggesting a possible resolution in that scenario.

3.2 Maturity Assessment

As per Fundamental Research project expectation, the three Solutions reached a current Maturity Level 1 (TRL-1). Table 2 describes for each criterion of TRL-1 the satisfaction of Solution 1, providing specific rationales. Table 3 refers to Solution 2. Table 4 refers to Solution 3.

Table 2. ER Fund / AO Research Maturity Assessment for MAHALO Solution 1

ID	Criteria	Satisfaction	Rationale - Link to deliverables - Comments
TRL-1.1	Has the ATM problem/challenge/need(s) that innovation would contribute to solve been identified? *- Where does the problem lie?* *- Has the ATM problem/challenge/need(s) been quantified that justify the research done? Note: an initial estimation is sufficient*	Achieved	The ATM community is struggling how to best utilise advances in ML/AI techniques in ways that keep the human ATCo at the centre of operations. MAHALO explored an operational concept where state-of-the-art ML-based automation served as an advisory system that either proposed optimal or personalised advisories and studied the impact on acceptance, agreement, trust and understanding.
TRL-1.2	Have the solutions (concepts/capabilities/methodologies) under research been defined and described?	Achieved	The ML modelling system under research and referring to MAHALO Solution 1 has been largely defined in its operational concept. In particular the interaction of its use with the User Interface, the whole Simulation Platform, the ML capabilities and the experimental procedures have been described. Each specific sub-feature of the operational concept is defined and described in a dedicated document. Those are: • State-of-the-art review of ML theory and design (D2.1). • Suitable Machine Learning (ML) capabilities, such as Supervised Leaning and Reinforcement Learning algorithms (D3.1 and D3.2).

			• Experimental user interface and ATC Conflict Detection and Resolution (CD&R) simulation capability (D4.1). • Conduction of validation trials of the user interface of the SectorX Simulator (D4.2). • Integration of ML capabilities with simulation platform (D5.1). • Developmental testing of integrated simulation capability (D5.2). • Specification of experimental design, for human-in-the-loop field study (D6.1). • Conduction and reporting on field study (D6.2).
TRL-1.3	Have assumptions applicable for the innovative concept/technology been documented?	Achieved	The ConOps of the Solution encompassed different assumptions: • Human and machine roles in a shared task of CD&R, during the interaction with an Advisory System. • Information requirements from the operator point of view in a shared task of CD&R, during the interaction with an Advisory System • Candidate ML architectures, such as Supervised Learning and Reinforcement Learning algorithms. • Candidate test protocols (e.g., data collection and analysis procedures).

			For further details, MAHALO initial and final ConOps versions are laid out in D2.2.
TRL-1.4	Have the research hypothesis been formulated and documented?	Achieved	The research hypotheses have been formulated. They detail what was the expected impact of the MAHALO ML (and transparency) approach on acceptance, agreement, trust and understanding. They are specified in D6.1 and D6.2.
TRL-1.5	Do the obtained results from the fundamental research activities suggest innovative solutions (e.g. concepts/methodologies/capabilities? - *What are these new concepts/methodologies/capabilities?* - *Can they be technically implemented?*	Achieved	The project suggested several innovative and technically feasible avenues for ML modelling, interface design and integration in controller working positions. They include: - Ecological approach to achieve interpretable ML models (WP4); - A pixel-based feature space observable by humans (i.e., Solution Space Diagram) (WP3); - Pipelines for data collection and ML learning (WP5); - Human interaction modes with ML advisories and integration into a controller working position (WP5); - ML approach for providing personalized and group-based resolution advisories. Further details about the above items are better described in D3.1, D4.1 and D5.1.

TRL-1.6	Have the potential strengths and benefits of the solution identified and assessed? *- Qualitative assessment on potential benefits. This will help orientate future validation activities.* *Optional: It may be that quantitative information already exists, in which case it should be used.*	Achieved	Field study results (WP6) defined simple single-group testing. Simulation and scenario context were judged to play a major role in observed results. Lessons were drawn from the fine-grained analysis, but broader lessons were also drawn for the need to design for personalisation in human-machine systems incorporating ML
TRL-1.7	Have the potential limitations, weaknesses and constraints of the solution under research been identified and assessed? *- The solution under research may be bound by certain constraints, such as time, geographical location, environment, cost of solutions or others.* *- Qualitative assessment on potential limitations. This will help orientate future validation activities.* *Optional: It may be that quantitative information already exists, in which case it may be used.*	Achieved	Related to TRL-1.6, one significant constraint in this research was the supply of sufficient data to permit ML training. Given logistical realities, it is difficult to collect enough training data to achieve a stabilised personal model (in this case, only 36 training samples were available per controller). This problem was mitigated by a synthetic process for creation of personal model advisories. Fortunately, the supply of data when pooled (across controllers) was sufficient to train the Supervised Learning ML to stability. In deliverable D6.2 additional limitations have been identified related to the CD&R scope and time constraints for administering experiment training, protocols and trials per participant.
TRL-1.8	Do fundamental research results show contribution to the Programme strategic objectives e.g. performance ambitions identified at the ATM MP Level?	Achieved	The impact ML will have on human performance in ATC had not been explored yet, even at a low TRL. The MAHALO project has tried to answer this question, in an attempt to identify a potential way to increase ATCO's acceptance, through the development of a system able to integrate automatic advisories based on ML and through

			two field simulations aiming at exploring the effect of machine conformance and transparency on human performance. MAHALO has now provided some initial results on ML in ATC. In deliverable D6.2 results and conclusions from post simulation data analysis are presented and discussed.
TRL-1.9	Have stakeholders been identified, consulted and involved in the assessment of the results? Has their feedback been documented in project deliverables? Have stakeholders shown their interest on the proposed solution?	Achieved	Via two workshops all relevant stakeholders, ranging from system developers at MUAC to controllers, have provided feedback and have expressed their interest in the proposed solution. The workshop results are made available on the MAHALO website.
TRL-1.10	Have initial scientific observations been communicated and disseminated (e.g., technical reports/journals/conference papers)?	Achieved	All the project observations and results have been communicated and disseminated. Below a list of dissemination activities. **Conference presentations and publications:** • Nunes Monteiro, T., Borst, C., Kampen, E. Van, Hilburn, B., & Westin, C. (2021). Human-interpretable Input for Machine Learning in Tactical Air Traffic Control. Proceedings of the Eleventh SESAR Innovation Days, 92, 1--6. + presentation (WP3 & WP4) • International Conference on Cognitive Aircraft Systems (ICCAS), 1-2 June 2022.

			Topics: Experimental design (WP6) and preliminary results (WP6). **Consortium and collaborator interchanges:** • TC2 Engage Workshop, 3 Sep 2021. Hilburn & Nunes. Topics: Developing ML capability (WP3); UI development (WP4) and ML/UI integration (WP5). • SJU ER4 Automation Workshop, 8 Mar 2021. Topics: MAHALO project overview, including goals, issues, and challenges.
TRL-1.11	Are recommendations for further scientific research documented?	Achieved	The results of data collection, ML modelling and real-time human interaction with the advisory system at different levels of conformance and transparency have been largely documented in D6.2 (Field simulation report). The document also addresses issues and topics relevant for further research. They include: - ML context sensitivity and the need to include contextual factors impacting CD&R performance (e.g., flight plans, time and separation targets, uncertainty, urgency, controller workload, etc.) - ML training challenges - Transparency challenges in tactical control settings

Table 3. ER Fund / AO Research Maturity Assessment for MAHALO Solution 2

ID	Criteria	Satisfaction	Rationale - Link to deliverables - Comments
TRL-1.1	Has the ATM problem/challenge/need(s) that innovation would contribute to solve been identified? *- Where does the problem lie?* *- Has the ATM problem/challenge/need(s) been quantified that justify the research done? Note: an initial estimation is sufficient*	Achieved	The lack of transparency functionalities in decision support systems and autonomous systems is a growing human factors concern. I ATC, this issue is envisioned to become increasingly important to solve with the introduction of digital assistants. MAHALO deliverable 2.1 (Integrated state of the art report) and deliverable 2.2 (Concept report) addressed this problem, discussed how transparency can help alleviate it, and explore previous approaches to transparency in research. The major consequence of opaque (non-transparent) in ATC is distrust and rejection of its contribution, leading to a potential loss in performance benefits.
TRL-1.2	Have the solutions (concepts/capabilities/methodologies) under research been defined and described?	Achieved	The concept, capabilities, methodology, and motivation for Solution 2 had been defined and described. Those elements refer in particular to: • Problem and motivation for solution (in D2.1) • Concept, capabilities, methodology (in D2.2) • Definition and description of SSD and agent transparency (in D4.1)

TRL-1.3	Have assumptions applicable for the innovative concept/technology been documented?	Achieved	The assumptions for Solution 2 was first documented in D2.2 (Concept report), describing the requirements for the transparency of proposed resolution advisories. Assumptions were finalised in D6.1 (Experimental design document).
TRL-1.4	Have the research hypothesis been formulated and documented?	Achieved	The research hypothesis related to Solution 2 matured over time as the project closed in on human-in-the-loop simulations. The various hypothesis documented were iteratively analysed: • Initial hypothesis and experimental plan (in D2.2) • First tests leading to revised experimental design (in D4.2) • Revised experimental plan and hypothesis (in D6.1) • Hypothesis answered in simulations (in D6.2)
TRL-1.5	Do the obtained results from the fundamental research activities suggest innovative solutions (e.g. concepts/methodologies/capabilities? *- What are these new concepts/methodologies/capabilities?* *- Can they be technically implemented?*	Achieved	Several innovative solutions leading up to Solution 2 has been discussed in deliverables D2.2, D3.1, D4.1 and D4.2. These were: - Transparency concepts for ML CD&R systems in ATC (D2.2, D4.1, D4.2); - Ecological approaches to achieve transparent ML models (D4.1); - Methodology for achieving a pixel-based feature space observable by

			humans (i.e., Solution Space Diagram) (D3.1); - Domain and agent transparency concepts and methodology (D6.1, D6.2). While domain transparency (i.e. SSD) was technically implemented, building on previous research at TUD, the content of the agent transparency concept (i.e. what to explain) was derived from manual analysis. However, the visual presentation of agent transparency was technically implemented in the interface.
TRL-1.6	Have the potential strengths and benefits of the solution identified and assessed? *- Qualitative assessment on potential benefits. This will help orientate future validation activities.* *Optional: It may be that quantitative information already exists, in which case it should be used.*	Achieved	Deliverable 6.2 (Field study report) documented the simulation results where Solution 2 was implemented. The hypothesis related to Solution 2, and effects of advisory transparency, were answered. The domain transparency level (SSD/diagram) was the preferred transparency level among most participants. However, quantitative results did not find a significant effect between transparency levels. The visual elements, driven by ecological interface design, supported participants in better understanding the separation distance that the advisory system was aiming for.
TRL-1.7	Have the potential limitations, weaknesses and constraints of the solution under research been identified and assessed?	Achieved	The constraints and limitations of Solution 2, as implemented in MAHALO, are briefly discussed in D6.2 (Field study repot). In the current implementation the Solution is bound to the

	- The solution under research may be bound by certain constraints, such as time, geographical location, environment, cost of solutions or others. *- Qualitative assessment on potential limitations. This will help orientate future validation activities.* *Optional: It may be that quantitative information already exists, in which case it may be used.*		SectorX simulation, but the concepts are transferable to any interface solution. The major limitation is the technical solution for agent transparency and have the ML advisory system generate the content (what to explain) to be visualised in the interface. Future research should explore ML interpretability models for deriving the content of the agent transparency solution.
TRL-1.8	Do fundamental research results show contribution to the Programme strategic objectives e.g. performance ambitions identified at the ATM MP Level?	Achieved	This is documented in D6.2 (Field study report) Solution 2 can contribute to improved performance and human-agent teamwork by conveying aspects of the ML agents underlying rationale for its advisories. The transparency mechanism that Solution 2 offers allows the user to derive a better understanding for why the advisory is presented and how it affects the overall situation. In turn, this facilitates better understanding of how the advisory affects the overall strategy that the human operator has for the CD&R task at hand. Note that this may actually drive rejection to the system, if the advisories goes against what the operator prefers. This knowledge – the potential discrepancy between operator preferences and system objectives provides valuable knowledge for either changing how the operator works (e.g. through training) or adjust the advisory system to conform with operator preferences.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

TRL-1.9	Have stakeholders been identified, consulted and involved in the assessment of the results? Has their feedback been documented in project deliverables? Have stakeholders shown their interest on the proposed solution?	Achieved	Key stakeholders and end users (i.e. air traffic controllers) were involved throughout the project and work related to Solution 2. Their feedback is documented in D4.2, D5.2, D6.2, D7.2 and D7.3.
TRL-1.10	Have initial scientific observations been communicated and disseminated (e.g., technical reports/journals/conference papers)?	Achieved	Initial findings related to Solution 2 has been disseminated at the ICCAS Toulouse, 1-2 june 2022. Submissions are in motion for the European Aeronautics Science Network (EASN), Towards Sustainable Aviation Summit (TSAS), and SESAR Innovation Days (SID).
TRL-1.11	Are recommendations for further scientific research documented?	Achieved	Recommendations for future research regarding Solution 2 and transparency is provided in this document (D7.4).

Table 4. ER Fund / AO Research Maturity Assessment for MAHALO Solution 3

ID	Criteria	Satisfaction	Rationale - Link to deliverables - Comments
TRL-1.1	Has the ATM problem/challenge/need(s) that innovation would contribute to solve been identified? *- Where does the problem lie?* *- Has the ATM problem/challenge/need(s) been quantified that justify the research done? Note: an initial estimation is sufficient*	Achieved	The ATM community is struggling how to best utilise advances in ML/AI techniques in ways that keep the human ATCo at the centre of operations. The MAHALO project explored ways how ML/AI systems could be integrated in ATC tasks and what their impacts are on controller acceptance, workload and system understanding. Several guidelines were then distilled based upon empirical insights obtained from the experiments, feedback from controllers and workshop results. The guidelines are divided in five categories: ML/AI design; Personalization; Transparency; HCI; and General.
TRL-1.2	Have the solutions (concepts/capabilities/methodologies) under research been defined and described?	Achieved	The several Guidelines distilled are the results and the outcomes of the whole technical work performed in the project, and take as input also the operational concepts deriving from Solutions 1 an 2. In particular the interaction of its use with the User Interface, ty refer to five cathegories (ML/AI design, Personalization, Transparency, HCI, General) and are described in D7.4 – Final Project Results Report.

Co-funded by the European Union

TRL-1.3	Have assumptions applicable for the innovative concept/technology been documented?	Achieved	The assumptions for Solution 3 are related to the technical work performed during the project and the results derived from the two Simulations carried out. Therefore, they were first documented in D2.2 (Concept report), describing the requirements for the transparency of proposed resolution advisories in Simulation fields. Assumptions were finalised in D6.1 (Experimental design document).
TRL-1.4	Have the research hypothesis been formulated and documented?	Achieved	Research hypotheses flowed out of an iterative experimental design, as documented primarily in deliverables D2.2, D4.2, and D6.1. Hypotheses were built up from research questions, and were stated as objective main- and interaction effect statements, to be tested in real time simulations.
TRL-1.5	Do the obtained results from the fundamental research activities suggest innovative solutions (e.g. concepts/methodologies/capabilities? - What are these new concepts/methodologies/capabilities? - Can they be technically implemented?	Achieved	MAHALO suggested several fruitful avenues for extending ML capabilities, and enhancing the human-ML collaboration in the enroute CD&R use case. They are documented in D6.2 and D7.4, and include: • The definition and expected impact advisory transparency and ML conformance on controller acceptance and agreement; • Redefinition of optimal ML to encompass controller workload, in addition to traffic geometric parameters; • The potential benefits of personalised advisory systems; and

			• Defining solutions in way that parallels operator (i.e. controller) goals.
TRL-1.6	Have the potential strengths and benefits of the solution identified and assessed? - Qualitative assessment on potential benefits. This will help orientate future validation activities. Optional: It may be that quantitative information already exists, in which case it should be used.	Achieved	Field study results (WP6) defied simple single-group testing. Simulation and scenario context were judged to play a major role in observed results. Lessons were drawn from the fine-grained analysis, but broader lessons were also drawn for the need to design for personalisation in human-machine systems incorporating ML.
TRL-1.7	Have the potential limitations, weaknesses and constraints of the solution under research been identified and assessed? - The solution under research may be bound by certain constraints, such as time, geographical location, environment, cost of solutions or others. - Qualitative assessment on potential limitations. This will help orientate future validation activities. Optional: It may be that quantitative information already exists, in which case it may be used.	Achieved	Related to TRL-1.6, one significant constraint in this research was the supply of sufficient data to permit ML training. Given logistical realities, it is difficult to collect enough training data to achieve a stabilised personal model (in this case, only 36 training samples were available per controller). This problem was mitigated by a synthetic process for creation of personal model advisories. Fortunately, the supply of data when pooled (across controllers) was sufficient to train the Supervised Learning ML to stability. Taking this constraint into account, the formulated list of Guidelines tries to highlight such a risk.
TRL-1.8	Do fundamental research results show contribution to the Programme strategic objectives e.g. performance ambitions identified at the ATM MP Level?	Achieved	The impact ML will have on human performance in ATC had not been explored yet, even at a low TRL. The MAHALO project has tried to answer this question, in an attempt to identify a potential way to increase ATCO's acceptance, through the development of a system able to integrate

			automatic advisories based on ML and through two field simulations aiming at exploring the effect of machine conformance and transparency on human performance. MAHALO has now provided some initial results on ML in ATC and the list of Guidelines builds up on this. Furthermore, in deliverable D6.2 specific results and conclusions from post simulation data analysis are presented and discussed.
TRL-1.9	Have stakeholders been identified, consulted and involved in the assessment of the results? Has their feedback been documented in project deliverables? Have stakeholders shown their interest on the proposed solution?	Achieved	ANSPs were involved throughout the project, both as research team members, and as simulation participants. Further, three ANSPs were represented on the two Workshops, and were consulted on both experimental design and results interpretation. Participating ANSPs have, at the individual controller level, shown interest in the MAHALO concepts broadly, and in the functioning of the advisory system. ANSP feedback is documented in D4.2, D5.2, D6.2, D7.2 and D7.3.
TRL-1.10	Have initial scientific observations been communicated and disseminated (e.g. technical reports/journals/conference papers)?	Achieved	Initial findings related to Solution 3 has been disseminated at the European Aeronautics Science Network (EASN) on October 20th 2022 and to EASA on October 21st 2022. A submission is also in motion for SESAR Innovation Days (SID).

TRL-1.11	Are recommendations for further scientific research documented?	Achieved	Recommendations for future research regarding Solution 3 and transparency is provided in this document (D7.4).

EUROPEAN PARTNERSHIP

Co-funded by the European Union

4 Conclusion and Lessons Learned

4.1 Conclusions

As already described in the Executive Summary, MAHALO intended to reach the following four main **Objectives**:

1. Develop an individually-tuned ML system comprised of layered deep learning and reinforcement models.
2. Couple this to an enhanced en-route CD&R prototype display to present machine rationale.
3. Evaluate in real-time simulations the relative impact of ML conformance, transparency, and traffic complexity, on controller understanding, trust, acceptance, workload, and performance.
4. Define a framework to guide design of future AI systems, including guidance on the effects of conformance, transparency, complexity, and non-nominal conditions.

These objectives are completely reached and accomplished, for the following reasons.

Objective 1: The project built a Machine Leaning algorithm, based on Supervised Leaning and able to provide personalised (conformal) resolution advisories. It was trained on resolution strategies data collected directly with Air Traffic Controllers.

Objective 2: MAHALO demonstrated that advisories presenting machine rationales could be integrated. It is also important to mention that for MAHALO Machine Learning explainability was addressed via a more operational approach, rather than what a literature on XAI would suggest using. MAHALO provided transparency, which has a broader content and definition in respect to explainability, via an ecological approach. It is important to note that this way to address explainability could be defined as “operational” instead of “developmental” and focuses more on the procedural and human factors aspects.

Objective 3: MAHALO conducted the following three human-in-the-loop (HITL) experiments:

- **Simulation 1:** The first simulation (performed in October 2021) used novices (e.g., university students). This was a developmental simulation, aimed at testing the fully integrated ML CD&R system, and its ability to provide conformal and transparency advisories. This was also the first experiment for testing the scenarios, data collection protocols, experiment procedures, questionnaire and debriefing materials, and data analysis procedures. Simulation 1 aimed at validating our simulation and analysis procedures, not at answering the MAHALO research hypotheses. These were addressed in simulations 2A and 2B.

- **Simulation 2A:** The second simulation hosted by DBL and ANACNA in Italy, involved 20 ATCOs as participants. The two phases (Conformance Pre-Test and Main Experiment) were performed

 Co-funded by the European Union

in early Dec 2021 and late Jan 2022. The same participants performed in both phases, with 19 returning for the Main Experiment.

- **Simulation 2B:** The third simulation, hosted by LFV Sweden, involved 16 ATCO participants and replicated the 2A simulation, with a minor adjustment to the simulator and a different controller cohort. The two phases (Conformance Pre-Test and Main Experiment) were performed in late Mar 2022 and late Apr 2022. The same participants performed in both phases, with 15 returning for the Main Experiment.

Objective 4: A list of Guidelines for future AI systems in ATC (Section 4.4) is described in this document and builds up on the results that have been derived from post experiment analysis.

To what concerns the Results derived from the technical work performed during the Project, the WP6 field study (see D6.2) data collection and analysis methodologies were directly driven by the Experimental design (D6.1), which identified several specific testable hypotheses regarding the relationship between conformance and transparency, and a number of dependent measures, primarily agreement, acceptance, and workload.

Analysis of conformance and transparency effects was challenged in the field study by the fact that scenario and simulation both emerged as extraneous variables that required separate 'fine-grained' analyses. That is, preliminary data analysis led the research team to reject a pooled data approach, and instead to treat combination of Simulation and Scenario as a separate sample. This obviously had the effect of reducing sample size and statistical power. Nonetheless, some clear effects and trends emerged.

4.1.1 Conformance effects

Recall that conformance here refers to the type of model (personal-, group-, or optimal) behind the advisory.

<u>Conformance and agreement</u>

For agreement, a statistically significant effect of conformance was found on agreement. In SIM2A, agreement was significantly lower for the group model in one scenario, and significantly lower for the optimal model another scenario. Conversely, SIM2A results showed a statistically significant higher agreement rating for the optimal model, in one of the two scenarios. These results underscore again the impact that simulation and scenario had as extraneous variables in the analysis.

<u>Conformance and acceptance</u>

Neither the pooled data nor the fine grained breakout data present a clear picture. Acceptance was very close across conformance levels. Acceptance (again, a five-level scale) was however consistently lower for the group condition. Recall that the acceptance data were not analysed using inferential statistics.

<u>Conformance and workload</u>

A main effect of conformance showed that, at least for one Simulation / Scenario combination, the personal model produced significantly lower workload ratings than did the optimal model. Although other workload effects failed to reach statistical significance, data plots suggest strong differences between the simulation sites.

4.1.2 Transparency effects

Transparency and agreement

In terms of advisory agreement, no main effects reached statistical significance. However, simulation and scenario effects were again apparent. Moreover, data trends suggested an interaction between conformance and transparency (which approached significance for SIM2A Scenario B). For the group model, the text condition showed the highest acceptance whereas for the optimal model, the vector condition showed the highest acceptance. One possible interpretation of these data is that a conformance by transparency interaction would suggest that in terms of controller acceptance, the most appropriate level of transparency display type might vary with the type of underlying conformance model.

Transparency and acceptance

It was hypothesised that controller acceptance of advisories would be higher if those advisories were presented in a high transparency display format. In terms of acceptance, descriptive data trends showed that whereas acceptance was very close across transparency levels, the text condition was associated with noticeably lower acceptance considering only full acceptance as the measure. This effect persisted but diminished as additional categories (nudge, adjust) were considered. Data trends also suggested the impact of simulation, scenario, and separation distance.

Transparency and workload

At one site, the vector condition showed a clear trend ($p>.05$) toward reduced workload, whereas at the other site the vector condition was associated with a reported workload increase.
One conclusion was that strong simulation site effects were likely influencing these results.

4.1.3 Self report and subjective feedback

The Field study report (D6.2) provides detailed results of the survey data collection. High level results of post-session questionnaire items can be summarised as follows:

- The vast majority of controllers found the ML solutions accurate, safe, easy-to-use, efficient, trustworthy, workload reducing, and helped speed up conflict resolution.
- Opinion was split on whether the system chose the same solution they would have.
- Controllers disagreed that system advisories were better than their own.

4.2 Lessons Learned

Over the course of the technical workflow of the various Work Packages, lessons were learnt with respect to design and training of ML models (WP3 Machine Learning model, and WP5 Integration), but also with respect to challenges in defining and measuring human interaction with advisory automation (WP4 Ecological User Interface, and WP6 Simulation), both in the current en-route CD&R use case, but also more broadly.

These lessons are discussed in the Field study report (D6.2), being the final outputs of the workflow converged in the final Work Package 6. They can be summarised as follows:

- Consistency of controller responses, within and between controllers, can impose constraints. If a given controller is internally inconsistent in his / her solutions over time, this complicates creation of a personalised model;
- Conversely, if controllers are externally inconsistent (i.e. there is disagreement between controllers) in their choice of solution, this makes it difficult to create a 'one-size-fits-all' conformal model;
- Given the observed and reported variability in controller solution strategies, and the 'large solution space' nature of the task (i.e., there is generally more than one workable solution strategy), It is not clear that en-route ATC CD&R is the ATM use case that can most benefit from ML. Other tasks, such as flow management, seem to have some traits (e.g. a smaller solution space) that might show more benefit;
- ML requires many training samples to stabilise learning. This was a known problem, but it was underscored by the challenges the MAHALO project faced in trying to create stable personal models based on a small number of individual training samples;
- Timing of solutions is critical—solutions that arrive after the controller has already devised a solution provide little benefit. Solutions that arrive too early can in fact represent an additional workload source (for the controller, who must now devise a solution, but also compare two solutions). This bias between 'early' and 'late' solvers seem to split the group in two;
- The definition of 'optimality' in advisories could refer to what is an optimal solution for a machine (in terms of traffic geometry), or for a human (in terms of preferences, physiological state, stress, fatigue, etc.). These two definitions can sometimes work at conflicting purposes;
- Context specificity, either in terms of simulation sample, chosen traffic scenario, or even solution preferences of a given controller (e.g., to prefer a tighter or looser CPA), play a large role in the development of ML advisory systems, and the idea that a one-size-fits-all approach would meet with widespread controller acceptance, is probably unrealistic.
- A large amount of data must be collected in order over a longer time period to facilitate ML generated personalized outputs. This requires an important work in organising the validation activities, trying to involve as much controllers as possible over a long time period. Their involvement should be planned at an early stage of the system design process, in order to consider their needs and input with the best human-centric approach.

It is important to note that MAHALO approach to use SSD as common feature to share between human and AI models might not, in hindsight, have resulted in the best possible outcome for the optimal AI models (e.g. RL). In other words, more optimised results could have been achieved when we used a different feature. We opted for the SSD to balance understanding/interpretability against optimality, but this balance might give suboptimal results.

4.3 Plan for next R&D phase (Next steps)

Results of the MAHALO project, while valuable in themselves, already suggest clear avenues of exploration for future research. The research team feels that the main thrust of the MAHALO results centres on how context and individual differences can drive the benefit of ML systems such as this.

One avenue that members of the team hope to explore is the potential utility of personalisation, or tuneable parameters that might allow for a hybrid of the optimal and personal model view. Such that controllers could tune certain parameters (and within a certain range) within the confines of an optimal advisory system. Separation margin appears to be the most prominent tuneable parameter to explore, based on results from the fine-grained analysis.

Second is the potential benefit of an adaptive systems approach, to actually automate that tuning such that parameters can be adjusted in step with learning performance. For example, the system might self-adapt the preferred solution timing or target CPA s controller experience grows.

Third is the potential hybridisation of ML together with adaptive systems, both for task performance but also as a trigger for task reallocation. That is, might ML potential benefits be improved if automation is given the authority to perform the task, but also the authority to shift task responsibility between human ad machine?

Finally, future research is required to explore potential benefits of advisory transparency on advisory acceptance and system trust in relation to ecological approaches, ML interpretability models, and the connection between the two. In contrast to expectations, measures of acceptance, agreement and workload did not benefit overall from increased transparency. To the contrary, increased transparency had in some Simulation and Scenario combinations an inverse effect with acceptance and agreement reaching higher values in the lowest transparency condition. This finding suggests that transparency alone may not be suitable as a measure for increasing operators' acceptance of advisories and trust in a system when that system performs different from the individual. By providing more information on why the system recommends a certain solution, the operator may become less willing to accept it. It is important to mention that this important result is pretty much aligned with those arising from other ER4 projects dealing with similar topics. It is a general thought that transparency and explainability could be potentially further explored in pre-tactical phases, where the planning operations are more critical and the need for information provided by a machine could increase.

MAHALO explored a narrow aspect of transparency – information about the degrees of freedom in the horizontal domain (i.e. domain transparency), and a text explanation supporting the visualized advisory with added information on separation distance. The concept of transparency I, however, a much broader construct and there are many aspects that require further research. This includes transparency mechanisms for supporting the ATCO in understand how the system works (e.g. the data processing, filtering, constraints etc in the model), how it derived a specific advisory (relationships between input data and output, and why the proposed advisory is considered best (e.g. best match to the individual, group, or optimized according to RL model).

4.4 Guidelines for future AI systems in ATC

The MAHALO project explored ways how ML/AI systems could be integrated in ATC tasks and what their impacts are on controller acceptance, workload and system understanding. Several guidelines were distilled based upon empirical insights obtained from the experiments, feedback from controllers and workshop results. The guidelines are divided in five categories:

1. ML/AI design
2. Personalization
3. Transparency
4. HCI
5. General.

Table 5. Guidelines for future AI systems in ATC: ML/AI Design

1	**ML/AI design**
ML/AI techniques can offer several benefits in finding solutions to traffic problems for which no analytical solution exist by considering multiple long-term (and sometimes competing) goals. Such ML-based optimisation, however, seems more appropriate for pre-tactical phases (e.g., multi-sector planning and airspace management), featuring a high degree of uncertainty, than for tactical operations in which controllers are faced with solving ad-hoc sector perturbations featuring relatively lower degrees of uncertainties. Additionally, when it is expected that humans need to collaborate with computerised agents capable of making decisions, it is often required that the system behaves consistent and is therefore predictable, as devised by the Human-Centred Automation (HCA) school of thought. ML solutions are generally governed by probabilities and therefore less predictable than conventional deterministic CD&R algorithms. Since the required amount of training data for ML/AI systems is often underestimated, a guideline is to design the ML/AI system at different levels of complexity, such that a fallback option is available when the highest complexity levels (in terms of number of states and actions) appear to be unfeasible with the limited amount of training data.	
1.1	Future AI systems for ATC should investigate which ML models that are best suited for balancing individual preferences and optimization approaches.
1.2	A large amount of data, collected over a longer time period, must be collected in order to facilitate ML generated personalized outputs.

Table 6. Guidelines for future AI systems in ATC: Personalisation

2	**Personalisation**
MAHALO and its predecessor MUFASA have demonstrated that personalisation helps in making an ML/AI-based advisory system more acceptable and easier to work with (e.g., faster response times). Personalisation can therefore be seen as a way to overcome some of the challenges associated with integrating ML/AI techniques in ATC from the perspective of human-machine collaboration. A prerequisite for personalisation in decision making is that sufficient intra-controller consistency and	

2	Personalisation
inter-controller variability exist in terms of actions/clearances. This requires a sufficiently large dataset to determine such existence before personalisation makes sense. MAHALO demonstrated that there is sufficient basis for personalisation in ATC decision making. The disadvantage of such personalisation is, however, that supervised ML/AI techniques aimed at modelling and mimicking an individual human controller could make the ATC system suboptimal. It is therefore recommended that the performance of personalised advisory systems is evaluated against target Key Performance Indicators, irrespective of sufficiently large intra-controller consistency and inter-controller variability.	
2.1	The development of future personalized AI systems for ATC requires end users' involvement in model development to ensure that the model captures what operators consider important for problem solving in the target task.
2.2	If ML models are to be trained on individual data, the model requires a lot of data from individuals to derive a solid and stable understanding of how that individual works, and what that individual's problem-solving preferences are. Model development should consider the use of synthetic (i.e. generated) data for training, to augment other data sources.
2.3	A suitable individual preference parameter for personalizing CD&R systems in conflict resolution choices is target separation distance.
2.4	Future ATC systems that are more personalized may lessen the need for them being transparent. A personalised system does not require high transparency, it reduces the need for transparency.
2.5	Individual preferences for parameters considered in conflict resolution decisions can be expected to vary between controllers. E.g. aircraft choice appears important for some but not others.
2.6	Controllers are more likely to accept and agree with a personalized system that adapts is recommendations to the individual's preferences.
2.7	There is no added benefit to acceptance or agreement of conflict resolution advisories that are shaped after the group of controller's preferences in terms of aircraft type, resolution direction, intervention time, and separation distance.
2.8	Future ATC system should explore personalization mechanisms to benefit system acceptance and agreement.
2.9	Future ATC systems should acknowledge and embrace in the design that controllers differ in their conflict resolution preferences.
2.10	Future ATC system should consider personalized applications when possible (i.e., taking into account a safety risk assessment).
2.11	Decision support systems capable of providing advisories/recommendations on actions should do so before the operator has made a decision on how to act (note that this can be before the action is implemented).

2	**Personalisation**
2.12	What aspects of a system that should be personalized should be driven by the operator's individual preferences in working and problem solving, and in what regards the operator is consistent over time.
2.13	Conformal (fully personalized) advisories should not be the main objective of future AI systems in ATC. That is, the system should not aim to only mimic human behaviour or decision making. Future systems should aim to optimize solutions but consider the individual operator's preferences and adjust the solution when feasible and appropriate. If the system goes against the individual's preferences, the system should be able to provide an explanation for why the system believes its solution to be better than the individual's.

Table 7. Guidelines for future AI systems in ATC: Transparency

3	**Transparency**
Similar to personalisation, transparency offers a way to increase the acceptance and understanding of (ML-based) advisory systems in ATC. MAHALO undertook an ecological approach in operationalising ML transparency by putting the emphasis on interpretable (visual) representations (here, Solution Space Diagram) rather than explainable ML models found in XAI fields. Controllers seemed to appreciate this approach as it puts ML solutions in the context of the problem that needs to be solved (i.e., traffic conflict). The same representation also served as a decision support tool, allowing controllers to formulate their own solutions and/or nudge the advised ML solution. This raises the discussion on what an operational controller might want and need to understand about the automated system and to what extent. For example, an ATCo might not need a deep insight into ML neural networks at the level of a ML system developer. Additionally, transparency needs are also affected by workload demands – in time-critical situations, an ATCo generally prefers to receive any workable solution and may not want to waste valuable cognitive resources in trying to understand that solution by digging through layers of information. In such cases, information on the resulting aircraft separation targeted by the advisory would be sufficient. Thus, given that transparency needs are likely context dependent and sensitive to operator preferences, we recommend an adaptive approach that allows ATCos to put machine decisions into context and lets them decide upon what they wish to see and when. Note that such an adaptive approach can be regarded as another form of personalisation, namely one that focuses on the preferred information that one wishes to see.	
3.1	Future AI system for ATC should focus on applying increased transparency for situations where the human and system work differently, and/or where the human have difficulties understanding the system. The need for transparency, and expected benefits, is higher for situations when system behaviour and advisories (e.g. it's output) is different from how the human operator prefers to work and solve problems. The need for transparency, and expected benefits, is higher for situations where the user does not understand system behaviour and advisories (e.g. it's output).

3	**Transparency**
3.2	Ecological interface design approaches can be used to increase the transparency of presented CD&R advisories by providing information on the constraints and solution possibilities affecting the control problem.
3.3	Future AI systems for ATC should investigate how transparency approaches can be used to improve system design. Increased transparency can benefit understanding of e.g. the system and/or situation, but does not necessarily benefit acceptance of a system and agreement with its advisories. The effect might be opposite, where increased transparency decreases acceptance and agreement.
3.4	Increased transparency supports better understanding or e.g., the system, its output, and/or the situation. With improved understanding, the operator can better determine if system behaviour or advisories are appropriate for the problem at hand. The use of the system partly depends on how the system's behaviour or advisory matches the preferences of the individual operator. As such, increased transparency can reveal to what extent the AI and operator work to solve problems similarly or different.
3.5	Transparency should ideally be individually tailored to facilitate a dialogue between humans and AI. Transparency should be provided in relation to what the user needs to understand, which requires the AI to be able to develop an understanding of the human it is interacting with.

Table 8. Guidelines for future AI systems in ATC: Human Computer Interaction

4	**Human Computer Interaction**
We believe that interaction flexibility is important for ATCo engagement. When humans are expected to play a central role in the system and bear the final responsibility over the safety of operations, human interaction with computerised systems is of paramount importance. That is, bearing responsibility without having authority is not the best position to be in. MAHALO showed one possible way of facilitating interaction by embedding it into existing controller tools. Via a conventional clearance menu, ATCos could not only accept, nudge, or change machine advisories, but also reject them and work with any other aircraft than the one receiving the advisory. Such flexibility was generally appreciated by ATCos as they felt empowered to influence the system in any way they preferred. We believe that offering such flexibility outweighs the (slight) performance decrements that could arise when ATCos change an optimal advisory into a suboptimal one. Note that affording flexibility in interaction can also be regarded as a form of personalisation.	

Table 9. Guidelines for future AI systems in ATC: General

5	**General**
Future ATC systems considering human-machines working together, should acknowledge that what is an optimal solution to a problem depend on the individual human (e.g., preferences, physiological state, stress, fatigue etc).	

5 References

5.1 Project Deliverables

[1] MAHALO D1.1, Project Management Plan, 30/06/2020

[2] MAHALO D2.1 Integrated SOAR, 01/01/2020

[3] MAHALO D2.2 Concept report, 03/12/2020

[4] MAHALO D3.1 ML report, 25/06/2021

[5] MAHALO D3.2 ML demonstrator, 06/08/2021

[6] MAHALO D4.1 E-UI design doc & demonstrator, 31/05/2021

[7] MAHALO D4.2 E-UI validation report, 31/05/2021

[8] MAHALO D5.1 Integration report, 13/12/2021

[9] MAHALO D5.2 Simulation 1 report, 13/12/2021

[10] MAHALO D6.1 Experimental design document, 15/10/2021

[11] MAHALO D6.2 Field simulation report, 06/06/202

[12] MAHALO D7.1 Project website, 02/07/2020, http://mahaloproject.eu/

[13] MAHALO D7.2 Workshop 1 report, 16/12/2021

[14] MAHALO D7.3 Workshop 2 report, 31/05/2022

[15] MAHALO D7.4 Final project results report, 20/06/2022

[16] MAHALO D7.5 Communication, Dissemination and Exploitation Plan, 16/10/2021

[17] MAHALO D8.1 PODP – Requirement No. 4, 30/07/2021

[18] MAHALO D8.2 PODP – Requirement No. 5, 31/05/2021

5.2 Project Publications

[19] Westin, C., Hilburn, B., Borst, C., Kampen, E. Van, Bang, M., (2020), Building Transparent and Personalized AI Support in Air Traffic Control, Digital Avionics Systems Conference, 2020

(Accessible here: http://mahaloproject.eu/wp-content/uploads/2020/10/DASC2020_Building-Transparent-and-Personalized-AI-Support-in-Air-Traffic-Control-FINAL.pdf)

[20] Westin, C., Nunes Monteiro, T., Borst, C., Bonelli, S., (2020), Balancing transparency and conformance of an AI conflict detection and resolution support system, Proceedings of the Tenth SESAR Innovation Days, 2020.

(Accessible here: http://mahaloproject.eu/wp-content/uploads/2021/01/MAHALO-SID2020-poster-description.pdf)

[21] Nunes Monteiro, T., Borst, C., Kampen, E. Van, Hilburn, B., & Westin, C. (2021), Human-interpretable Input for Machine Learning in Tactical Air Traffic Control. Proceedings of the Eleventh SESAR Innovation Days, 92, 1--6. + presentation (WP3 & WP4).

(Accessible here:

http://mahaloproject.eu/wp-content/uploads/2022/06/SID_Paper_2021_MAHALO-2.pdf)

[22] Bonelli, S., Borst, C., Brambati, F., Cocchioni, M., Hilburn, B. & Monteiro-Nunes, T., (2022), Transparent and Personalised AI Support in Air Traffic Control: First Empirical Results, The International Conference on Cognitive Aircraft Systems, 2022.

(Accessible here:

https://drive.google.com/file/d/185A19xMxsohTgnn2u3sFjllFxIqf8lXo/view?usp=share_link)

5.2.1 Further Project Publications

[23] EASN article publication, November 2022

[24] SID article publication, December 2022

Appendix A

A.1 Acronyms and Terminology

Table 10. Acronyms and Terminology

Term	Definition
AI	Artificial Intelligence
ANN	Artificial Neural Network
APP	Approach Control
ATC	Air Traffic Control
ATM	Air Traffic Management
CD	Conflict Detection
CD&R	Conflict Detection and Resolution
CNN	Convolutional Neural Network
CPA	Closest Point of Approach
CR	Conflict Resolution
CWA	Cognitive Work Analysis
DDPG	Deep Deterministic Policy Gradient
DQfD	Deep Q-learning from Demonstrations
DQN	Deep Q-Network
DRL	Deep Reinforcement Learning
E-UI	Ecological User Interface
EID	Ecological Interface design
FFNN	Feed-Forward Neural Network
GBM	Gradient Boosting Machines
KNN	K-Nearest Neighbour
LOA	Level/s of Automation

LSTM	Long Short-Term Memory
MAHALO	Modernising ATM via Human-Automation Learning Optimisation
MDP	Markov Decision Process
ML	Machine Learning
MLR	Multiple Linear Regression
MSAW	Minimum Safe Altitude Warning
MTCD	Medium-Term Conflict Detection
MUFASA	Multidimensional Framework for Advanced SESAR Automation
NN	Neural Network
RF	Random Forest
RL	Reinforcement learning
S3JU	SESAR3 Joint Undertaking (Agency of the European Commission)
SA	Situation Awareness
SL	Supervised Learning
SOAR	State of the Art Report
STCA	Short-Term Conflict Alert
SESAR	Single European Sky ATM Research Programme
SVM	Support Vector Machine
SSD	Solution Space Diagram
TCT	Tactical Controller Tool
UI	User Interface
UL/USL	Unsupervised Learning