D6.1 – Report on the Validation Plan-Exploratory Research Plan (ERP)

Deliverable ID:	**D6.1**
Project acronym:	**TRUSTY**
Grant:	**101114838**
Call:	**HORIZON-SESAR-2022-DES-ER-01**
Topic:	**HORIZON-SESAR-2022-DES-ER-01-WA1-7**
Consortium coordinator:	**MÄLARDALEN UNIVERSITY**
Edition date:	**20 December 2024**
Template edition:	**02.00.01**
Edition:	**00.03.00**
Status:	**Official**
Classification:	**PU**

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Abstract

The main objective of this document is to present the Exploratory Research Plan (ERP) for the TRUSTY project. This document details the TRUSTY concept as an AI-based tool for assistance in Air Traffic Control Service delivery in the Remote Digital Tower environment. The details herein describe the use of the TRUSTY tool for taxiway and runway monitoring, and decision support, whilst optimising the cognitive demands on the Air Traffic Control operator. This plan also describes in detail the validation exercise which will take place to elicit information and data relating to the performance levels being achieved by the TRUSTY concept. Scientifically rigorous experimental design methodologies have been applied to the planning of the validation exercise, to ensure unbiased and well-controlled data generation. This is for the purpose of demonstrating proof of the TRUSTY concept and its performance contribution to improvements in the ATM environment.

Authoring & approval

Author(s) of the document

Organisation name	Date
Deep Blue	18/12/2024
ENAC	16/07/2024

Reviewed by

Organisation name	Date
MDU	18/12/2024
ENAC	18/12/2024
UNIROMA1	18/12/2024
Deep Blue	18/12/2024

Approved for submission to the SESAR 3 JU by

Organisation name	Date
MDU	18/12/2024
ENAC	18/12/2024
UNIROMA1	18/12/2024
Deep Blue	18/12/2024

Rejected by[1]

Organisation name	Date

Document history

Edition	Date	Status	Organisation author	Justification
00.00.01	10/05/2024	Draft	Deep Blue	First draft produced
00.00.02	28/05/2024	Draft	Deep Blue	Follow on draft
00.00.03	16/06/2024	Draft	ENAC	Add content to Sections
00.00.04	20/06/2024	Draft	Deep Blue	Draft for the first review
00.00.05	26/06/2024	Draft	Deep Blue	First review considered

[1] Representatives of the beneficiaries involved in the project.

00.00.06	27/07/2024	Draft	MDU	Make a version for SJU
00.01.00	04/07/2024	Submitted as Draft	MDU	Submitted to SJU
00.01.01	24/07/2024	Draft	MDU	Updated based on the Review of the report
00.01.02	29/07/2024	Draft	UNIROMA1	Updated based on the Review of the report
00.01.03	29/07/2024	Draft	ENAC	Updated based on the Review of the report
00.01.04	16/08/2024	Draft	Deep Blue	Address SJU review comments
00.01.05	18/08/2024	Draft	Deep Blue	Final complete Version for submission
00.01.06	20/08/2024	Draft	MDU	Final review and make ready for SJU submission
00.02.00	20/08/2024	Final 1st Round	MDU	Ready for submission
00.02.01	22/09/2024	Draft	Deep Blue	Addressed SJU comments
00.02.02	26/09/2024	Draft	Deep Blue	Sent for internal review
00.02.03	26/09/2024	Draft	MDU	Review of the document
00.02.03	03/10/2024	Draft	UNIROMA1	Review of the document
00.02.03	04/10/2024	Draft	ENAC	Review of the document
00.02.04	07/10/2024	Draft	Deep Blue	Ready for submission
00.02.00	07/10/2024	Final 2nd Round	MDU	Submitted
00.02.00	17/12/2024	Draft	SJU	Received SJU comments
00.02.01	18/12/2024	Draft	Deep Blue	Addressed SJU comments
00.02.02	19/12/2024	Draft	MDU	Update table 8
00.03.00	20/12/2024	Final	MDU	Ready for submission

Disclaimer

The opinions expressed herein reflect the author's view only. Under no circumstances shall the SESAR 3 Joint Undertaking be responsible for any use that may be made of the information contained herein.

TRUSTY

TRUSTWORTHY INTELLIGENT SYSTEM FOR REMOTE DIGITAL TOWER

TRUSTY

This document is part of a project that has received funding from the SESAR 3 Joint Undertaking under grant agreement No 101114838 under European Union's Horizon Europe research and innovation programme.

Table of contents

List of figures

List of tables

Executive summary

This document describes the approach, validation objectives and methods that are planned for performing validation activities in the SESAR-supported TRUSTY project.

The purpose of the validation exercise described in this plan is to explore the integration of Artificial Intelligence into Air Traffic Management, specifically in the Remote Digital Tower environment and its impact on Human Performance associated with trust in technology. This is achieved through situating the concept in the context of several research and innovation needs. These needs have been identified regarding the strategic trajectory of research and development in the field of air traffic control, enhanced automation, novel technologies and increased flexibility in the provision of an Air Traffic Management application domain. The TRUSTY project aims to contribute to a greater understanding and insight in these research areas and so a potential solution has been proposed, as a conduit for greater exploration, in addition to the demonstration of a proof of concept.

This plan describes in detail the validation exercise which has been designed to elicit information and data relating to the performance levels being achieved by the TRUSTY concept. This exercise has been designed around several validation objectives and success criteria, which themselves, distil identified research and innovation needs into something observable and measurable. Scientifically rigorous experimental design methodologies have been applied to the planning of this exercise, to ensure unbiased and well-controlled data generation. Human Performance validation objectives have been generated through the use of the Human Performance Assessment Process.

The overall aim of this exploratory research plan is to guide the development and evaluation of the TRUSTY concept and to progress the TRUSTY project and solution from its inception at TRL0 to TRL1.

This will be achieved through the following objectives:

- To further define the tactical functioning of the TRUSTY solution;
- To investigate the potential performance impacts of the TRUSTY solution;
- To generate further insight into the concept of ‘trust’ during the integration of an AI solution in ATC/ATM;
- To explore the benefits of AI in the Remote Digital Tower environment.

The validation exercise described assesses the concept of the TRUSTY solution in a real-time simulator exercise, using a representative flight simulator facility and employment of operational ATCOs in the test of the technology in various scenarios. A validation activity will occur in the simulator facilities at the Ecole Nationale de l'Aviation Civile (ENAC) Toulouse, where access to professional ATC operators can be guaranteed.

The validation scenarios for the TRUSTY project are designed to be consistent with the operational concept of a dual-task scenario involving ATM provision at an in-situ airfield and airfield supervision at a remote airfield. The validation exercise includes both a reference scenario (baseline) without the TRUSTY solution, and a scenario with the TRUSTY solution integrated into the RDT displays.

1 Introduction

1.1 Purpose of the document

The purpose of this deliverable, D6.1 'Report on the Validation Plan', is to provide an Exploratory Research Plan (ERP) for the project, which describes the development of the TRUSTY solution from its inception at Technical Readiness Level (TRL) 0. This document will present details of the solution and then discuss the intended progress required to mature it to TRL1. This progress will involve situating the concept in the context of several research and innovation needs, from which research questions arise. These needs have been identified regarding the strategic trajectory of research and development in the field of air traffic control, enhanced automation, novel technologies and increased flexibility in the provision of an Air Traffic Management (ATM) application domain. The TRUSTY project aims to contribute to a greater understanding and insight in these research areas and so a potential solution has been proposed, as a conduit for greater exploration, in addition to the demonstration of a proof of concept.

To measure the TRUSTY projects' performance contribution to improvements in the ATM environment, an extensive plan has been presented in this ERP. The plan highlights the areas of contribution that the TRUSTY project can hypothetically make. This is in accordance with Key Performance Areas highlighted in the strategy planning of the European Commission, articulated in the SESAR Performance Framework [6]. Within this strategic planning and more broadly in ATM modernisation, it is apparent that greater exploration of human performance whilst using technically advanced solutions is necessary, especially in the field of AI. Therefore, the TRUSTY concept is focussed on the very distinct area of trust in AI, which is a critical area for ensuring the optimal integration of the human and AI, into important and safety-critical areas of ATM.

This plan also describes in detail the validation exercise which has been designed to elicit information and data relating to the performance levels being achieved by the TRUSTY concept. This exercise has been designed around several validation objectives and success criteria, which themselves, distil identified research and innovation needs into something observable and measurable. Scientifically rigorous experimental design methodologies have been applied to the planning of this exercise, to ensure unbiased and well-controlled data generation.

The purpose of this document therefore is to present a plan that will be used to progress the TRUSTY project and solution from its inception to TRL1. This will be achieved through the strict application of scientifically sound and validated methodologies, in generating authentic and accurate data, which will demonstrate proof of the TRUSTY concept and its performance impact.

1.2 Intended readership

The intended readership of this document is the community whose aspiration is to advance knowledge in the design, development and integration of novel technologies in the air traffic management environment and other safety critical areas. Therefore, individuals from industry, academia, regulatory bodies, aviation professionals and advisory consultants, who have an interest in taking advantage of new technologies to improve the safety and efficiency of the aviation sector. As this document includes the description of the validation exercise for the TRUSTY project, its contents are also relevant to the research community in terms of experimental design and the application of scientific methodology.

Moreover, as the TRUSTY solution proposed in this project is in its infancy, this document should demonstrate to all named communities the possible employment of AI in the ATM community and will hopefully inform and inspire others in the exploration of this field.

1.3 Background

The TRUSTY project builds on the work conducted in the ARTIMATION project, a SESAR JU funded research project, extending its scope to a more comprehensive exploration of trustworthiness in AI systems. TRUSTY is dedicated to studying AI trustworthiness from both algorithmic and human perspectives, leveraging the insights gained from ARTIMATION.

ARTIMATION focused on the impact of AI explanations on user behaviour, cognition and acceptance, with a particular focus on differences in individuals' expertise. The main findings from the ARTIMATION project, which have had a direct impact on the research objectives in the TRUSTY project, were those associated with the impact of explainability on cognitive workload. This is because Explainable AI has enabled users to concentrate on critical aspects without needing to understand the AI's internal workings. TRUSTY aims to further investigate how varying levels of explainability can optimize cognitive workload management in relation to aspects of trust. Another important finding from ARTIMATION was that of the perspective of users, according to their age and experience. This was in relation to their willingness to adopt future digital assistance, wherein the lesser experienced users perceived a greater impact on daily operations than expert users, underscoring the need for tailoring AI explanations not only in a technical sense, but also according to the technological expertise of the user. Additionally, ARTIMATION found that users often believed that the AI was 100% accurate. This bias indicated the need to investigate the trustworthiness of AI further, and thus the TRUSTY project was established [9].

Beyond ARTIMATION, other research projects funded by the SESAR JU are summarised in Table 1. Relevant research is reported in [11], where advancement was made in the integration of AI to detect conflicts in air traffic by analysing aircraft surveillance data, thereby augmenting situational awareness for controllers. Guidance on updating the regulatory framework for explainability, machine learning systems in ATMs, transparency and user acceptance has been reported in [12]. In [13], the authors discussed how AI can play a crucial role in enhancing the resilience of ATM operations against disruptions caused by reliance on network infrastructures and remote sensors. In [14] the authors explored how blockchain technology and self-learning networking architectures can be integrated with explainable AI to build trust in human agents and optimise air traffic control. In [15] the authors presented work on the introduction of digital ATCOs, capable of autonomously performing time-consuming tasks, emphasising the importance of a human-autonomy teaming interface, supported by explainable AI. These advancements contribute, alongside the SESAR JU supported projects to the introduction of AI-driven decision-making in ATM/ATC, bringing more flexibility, transparency, reliability, and acceptability to human operators.

Table 1 Summary of relevant SESAR JU funded research projects contributing to the development of AI in ATM

Project Name	Timeline	Contribution
AISA (https://doi.org/10.3030/892618)	June 2020 – Nov 2022	Strategy for providing the necessary information to a specific ATM operational environment (en-route ATC) to make them trust the automated system.

MAHALO (https://doi.org/10.3030/892970)	June 2020 – Nov 2022	AI-based Conflict Detection & Resolution tool with different levels of conformance and transparency.
TAPAS (https://doi.org/10.3030/892358)	June 2020 – Nov 2022	XAI methods for two operational cases: Conflict Detection & Resolution applied to ATC (tactical), and Air Traffic Flow Management (ATFM) (pre-tactical).
ARTIMATION (https://doi.org/10.3030/894238)	Jan 2021 – Dec 2022	Tools for Conflict Detection & Resolution and Delay Prediction with explanation through visualisations.
SAFEOPS (https://doi.org/10.3030/892919)	Jan 2021 – Dec 2022	A decision-support tool powered by AI to help ATCOs make complex decisions in the context of go-arounds.

1.4 Structure of the document

This document commences with a description of the TRUSTY concept and outlines the challenge that the TRUSTY project aims to address in the 'Problem statement'. In these early sections, the operational context and key scenarios have been described. This is to engender a common understanding of what the TRUSTY concept is.

Proceeding this are sections which present the basis for the research, including the aims, objectives, experimental environment, assumptions and scope. In the same section, the key Research and Innovation (R&I) needs are described, which have associated with them research questions in need of exploration. In the mid sections of the document, estimations have been presented on the potential performance contributions of the TRUSTY concept, in relation to strategic Key Performance Areas. Also, at this point there is a description of the route that the TRUSTY project will take to progress from its initial to its exit maturity level, in terms of TRL.

In the latter parts of the research plan, the research approach is presented. This will demonstrate how the performance of the TRUSTY solution will be measured, starting from a set of validation objectives and success criteria. These validations objectives encompass all the areas that the project will measure performance against. In the final part of the document, there is a description of the validation exercise, which has been designed in such a way as to collect data and evidence to show performance against the validation objectives.

1.5 Glossary of terms

Table 2 Glossary of terms

Term	Definition	Source of the definition
Air Traffic	All aircraft in flight or operating in the manoeuvring area of an aerodrome.	ICAO Annex 11 - ATS
Air Traffic Controller	Qualified following ICAO Annex 1 – Personnel Licensing and holding a rating appropriate to the assigned functions, A person authorized to provide air traffic control services.	EUROCONTROL ATM Lexicon
Air Traffic Management	The dynamic, integrated management of air traffic and airspace including air traffic services, airspace management and air traffic flow management – safely, economically, and sufficiently – through the provision of facilities and seamless services in collaboration with all parties and involving airborne and ground-based functions.	ICAO 4444 - ATM
Air Traffic Services	A generic term meaning various, Flight Information Service (FIS), Alerting Service (ALRS), and Air Traffic Control Service (ATC) (area control service, approach control service, or aerodrome control service). In this document, when the term ATS is used, it is usually referring to TWR or AFIS.	ICAO, Annex 11
Artificial intelligence (AI)	Technology that can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with.	[19]
Conventional tower	A facility located at an aerodrome from which aerodrome ATS is provided principally through direct out-of-the-window observation of the aerodrome and its vicinity;	[18]
Remote Tower	A geographically independent facility from which aerodrome ATS is provided principally through indirect observation of the aerodrome and its vicinity, employing a visual surveillance system.	[18]
Single mode of operation	The provision of ATS from one remote tower/remote tower module for one aerodrome at a time.	[18]
Multiple modes of operation	The provision of ATS from one remote tower/remote tower module for two or more aerodromes at the same time (i.e. simultaneously).	[18]
Trustworthy AI	An AI system which is valid and reliable, safe, secure and resilient, accountable and transparent, explainable and	

	interpretable, privacy-enhanced, and fair with harmful bias managed	

1.6 List of acronyms

Table 3 List of acronyms

Term	Definition
ACHIL	Aeronautical Computer Human Interaction Laboratory
AI	Artificial Intelligence
ANOVA	Analysis of Variance
APSARA	AI-Powered Situational Awareness for Remote Airfields
ATC	Air Traffic Control
ATCO	Air Traffic Control Officer
ATM	Air Traffic Management
ATS	Air Traffic Services
BLE	Blue Tooth low-energy
CDE	Communications, Dissemination and Exploitation
CPDLC	Controller Pilot Data Link Communications
CTAF	Common Traffic Advisory Frequencies
DES	Digital European Sky
EDA	Electrodermal Activity
EEG	Electroencephalogram
ERP	Exploratory Research Plan
ERR	Exploratory Research Report
FOD	Foreign object damage
GA	Grant Agreement
GDPR	General Data Protection Regulation
GSR	Galvanic Skin Response
GUI	Graphical User Interface

HAIT	Human-AI Teaming
HCE	Human-Centred Explainable
HE	Horizon Europe
HF	Human Factors
HMI	Human Machine Interface
HP	Human Performance
HPE	Human Performance Envelope
ID	Identifier
ISA	Instantaneous Self-Assessment
ISO	Impact and Societal Objectives
KPA	Key Performance Area
KPI	Key Performance Indicator
LFBO	Toulouse-Blagnac international airport
LFBR	Muret-Lherm Aerodrome
LIDAR	Light detection and ranging
MAWP	Multiannual work programme
ML	Machine Learning
MML	Multimodal Machine Learning
OBJ	Objective
OPEFF	Operational effectiveness
pBCI	passive Brain Computer Interface
PU	Public
PI	Performance Indicator
RDT	Remote Digital Tower
RTS	Real Time Simulation
SA	Situational Awareness
SART	Situational Awareness rating technique

SAF	Safety
SC	Success Criteria
SESAR	Single European Sky ATM research
SESAR 3 JU	SESAR 3 Joint Undertaking
SIO	Scientific, Innovation, And Research Objectives
SME	Subject Matter Expert
SRIA	Strategic research and innovation agenda
SVG	Scalable Vector Graphics
TMA	Terminal Manoeuvring Area
TRL	Technical Readiness Level
TRUSTY	Trustworthy Intelligent System for remote digital tower
UC	Use Case
UCI	User Centric Interface
UX	User Experience
VHF	Very High Frequency
WL	Workload
XAI	Explainable Artificial Intelligence

2 Concept outline

2.1 Problem statement

Remote Digital Towers (RDTs) are already in use in many parts of the world, providing an ATM service for remote airfields, at which the ATCO is not co-located at the specific airfield. RDTs largely rely on video presentation and remote sensors to provide the ATCO with the information needed to provide the ATC service. Due to safety concerns and technology limitations at remote airfields, data processing and monitoring are today mainly done by the ATCO. Such monitoring is time-consuming and can generate low vigilance and fatigue when events e.g. aircraft movements, are infrequent at remotely managed airfields. Alternatively, this same situation can produce considerably high workload with increased potential for error, if the traffic is too dense or too complex. The RDT approach could benefit from enhanced automation for data processing, to assist the ATC in assimilating the information required for accurate situational awareness, efficient decision making and WL optimisation.

Moreover, because of the challenge from safety concerns in this field, the requirements for reliability, transparency, accountability, and user trust and acceptance, are high with the introduction of any novel technologies. For this reason, the TRUSTY project is focused specifically on exploring user confidence in an AI system in which acceptability and trust are maximised. These are maximised by using transparent and explainable algorithms combined with the optimal presentation of AI outputs on an appropriately designed user interface.

2.2 Concept description and operational scenarios

In TRUSTY, SESAR researchers are investigating and validating an AI candidate solution for the RDT environment. The AI solution supports the operator in runway and taxiway monitoring, through object detection and characterisation, providing the ATCO with an increased awareness of runway and taxiway incursions, or other changes in the situation at the remote airfield e.g. windsock orientation. Moreover, the project will consider the use of multi-modal inputs e.g. video, audio, and communications data, to strengthen the AI output. This data is processed through the AI to detect situations with increased risk and, through the use of an effectively designed interface, to bring the attention of the remote ATCO to these situations. The TRUSTY solution thus reduces the demand on the operator in the task of airfield monitoring, improving situational awareness and increasing the capacity of the ATCO to safely attend to multiple tasks.

2.2.1 Operational/Technical context

The TRUSTY solution operates within the RDT environment, utilizing advanced digital technologies to provide ATC services remotely. The operational environment is concentrated at the controller workstation, equipped with high-resolution video displays, sensor data feeds, and an integrated communication interface. As previously mentioned in Report D3.2 [16], the operational context of the TRUSTY project is situated in the concept of the dual operating mode, as established by EASA [18], meaning that the controller will have to provide ATS to two airports simultaneously from a single tower facility, consequently monitoring multi-environments. Within the TRUSTY project this operating mode comprises a simulated in-situ tower environment, for conducting real ATC tasks within a conventional

airport environment, namely Toulouse-Blagnac international airport (LFBO), and a second nearby airfield, operating as a remotely managed airfield, namely Muret-Lherm Aerodrome (LFBR).

Beyond the ATC duties at the conventional tower, the controller will therefore be required to execute a supervisory role as a RDT ATCO, at a workstation where information monitors will display real-time video feeds, which originate from cameras situated around the remote airfield. These cameras offer comprehensive visual coverage of both runways and taxiways. Additionally, the workstations will have the potential to integrate data from remote sensors such as radar, weather sensors, and audio inputs, providing a holistic view of airfield conditions.

The TRUSTY AI algorithms are aimed at analysing this data to enhance situational awareness and support decision-making by identifying relevant patterns and anomalies, such as potential runway incursions or adverse weather conditions. The AI system will be designed to be robust and adaptable, ensuring it can handle varying environmental conditions whilst providing accurate and timely alerts. The AI outputs will be presented on a user-centric interface, delivering clear and actionable information with visual and auditory alerts to guide controllers' attention to critical situations. A key goal of the TRUSTY project is to assess how trust in the system fluctuates according to the transparency and explainability of the AI, as presented in interface visualisations. By providing detailed situational analysis, the TRUSTY solution aims to reduce cognitive demands for controllers, enabling them to focus more effectively on complex decisions without causing fatigue or error.

2.2.2 Key scenarios

The operational context of TRUSTY includes a set of use-case scenarios designed to assess the ATCO's trust in the solution within the dual operating mode described above. These scenarios, originating from insights gleaned from project workshops and previous analyses [16], will evaluate how ATCOs interact with the XAI system during cognitively demanding situations. The scenarios involve tasks which cover conventional ATC duties, such as take-off/landing clearances and ground movement coordination, as well as tasks related to the remote airfield e.g. monitoring events through high-definition video feeds and sensor data.

Two of the key scenarios for the TRUSTY project are described as follows:

Scenario 1: Enhanced Runway and Taxiway Monitoring

Context: In an RDT environment, ATCOs are required to manage multiple remote airfields simultaneously. This includes monitoring runway and taxiway conditions to ensure safe and efficient aircraft movements.

Description: The TRUSTY project's XAI system is integrated into the RDT setup, leveraging multi-modal inputs such as video feeds and audio communications. The XAI system continuously processes this data to detect anomalies, such as unauthorized vehicles, debris, or adverse weather conditions (e.g., intense fogs, wind shear) that could affect runway and taxiway operations.

Operational Flow:

> *Detection*: The XAI system identifies an unauthorized vehicle entering the runway at Airfield A. The system flags this anomaly and generates an alert.

Notification: The alert is presented on the ATCO's user interface with a clear signal, highlighting the exact location and nature of the intrusion. Here, the explainability and the accuracy of the system allow for studying its impact on the trust of the user.

Action: Relying on the XAI's rationale through the explainable AI interface, the ATCO coordinates with ground services in situations involving vehicles or ground obstacles. In the case of an aircraft on the runway, the ATCO utilizes the same AI guidance but follows established protocols, directly communicating with the involved aircraft, and relevant ATC units, and approaching aircraft to resolve the issue effectively.

Outcome: The timely detection and clear communication prevent a runway incursion, maintaining safety and operational efficiency.

Impact: This scenario demonstrates the capability of the TRUSTY XAI system to enhance situational awareness and decision-making, reducing cognitive load and improving safety in complex, multi-airfield operations.

Currently the runway incursion detection systems are deployed in the hot spots of the airports through which area most of the aircraft and vehicles pass. According to [17], the prevailing systems are based on video and infrared imaging position sensors, infrared radiation position sensors, induction coil position sensors, leaky cable position sensors, microwave radiation position sensors, regional microwave detection position sensors, etc. In summary, all the prevailing systems include different sensors to detect the runway incursions. To the best of our knowledge, no prevailing runway incursion system deploys object detection technology with an explanation for live video feeds. However, the proposed system in the TRUSTY project is based on high-definition video cameras where object detection technologies are implemented on live video feeds with explanations at different levels on each instance of object detection.

Scenario 2: Adaptive Weather Management

Context: Weather conditions significantly impact air traffic operations, necessitating real-time monitoring and adaptive management to ensure safety. In a RDT setup, ATCOs rely on accurate weather predictions and clear communication of potential impacts on operations.

Description: The TRUSTY AI system integrates a communications detection system between pilots and ATCOs, to manage sudden weather changes and their effects on flight operations.

Operational Flow:

Monitoring: Pilots communicate an observed sudden wind shift to ATCOs during routine radio checks. The XAI system analyses these communications to understand the weather change's potential impact.

Alert Generation: Based on pilot reports and XAI analysis, an alert is generated, explaining the predicted wind shift's impact on aircraft performance and safety. Here, the explainability and the accuracy of the system allow for studying its impact on the trust of the user.

Coordination: The ATCO receives the alert with a clear explanation, allowing them to relay precise instructions to pilots regarding adjusted take-off and landing protocols.

Simultaneously, the XAI system can assist in analysing communications between pilots to ensure all are aware of the new conditions. These communications can be retrieved by monitoring VHF frequencies and keeping track of CTAF, Multicom, and 123.450 MHz for air-to-air communications for example.

Outcome: Proactive adjustments to flight operations based on pilot reports and XAI-coordinated communications prevent potential incidents, ensuring safety and maintaining operational efficiency.

Impact: This scenario demonstrates the capability of the TRUSTY XAI system to enhance situational awareness and decision-making by highlighting the importance of explainable AI in enhancing trust and reliability in weather management through effective communication.

3 Context of the exploratory research plan

3.1 Exploratory research plan context

The overall aim of this exploratory research plan is to guide the development and evaluation of the TRUSTY concept.

This will be achieved through the following objectives:

- To further define the tactical functioning of the TRUSTY solution;
- To investigate the potential performance impacts of the TRUSTY solution;
- To generate further insight into the concept of 'trust' during the integration of an AI solution in ATC/ATM;
- To explore the utility of AI in the Remote Digital Tower environment.

The validation activity will occur in the simulator facilities at the Ecole Nationale de l'Aviation Civile (ENAC) Toulouse, where access to professional ATC operators can be guaranteed. Any bench testing of the AI solution will occur at the place of development, namely Mälardalen University. Moreover, any pre-trial workshops or questionnaires will occur online. Deep Blue assumes the role of Validation Lead and as such is responsible for leading validation planning and management. ENAC will take the responsibility of validation conduct and data generation with the support of consortium partners. La Sapienza University of Rome will lead the neurophysiological assessment.

This detail will be further expanded upon, in the validation exercise planning later in the document.

3.2 Scope

This document describes the developmental plan for the TRUSTY solution which has been previously detailed in the concept outline in section 2. It will cover the approach to performance assessment and validation exercise planning in the project, within an operational context. This document however does not address the design of the AI or the design of the user interface, these activities are being undertaken in Work Packages 4 'Trusted Intelligent System Development' and WP5 'AI-Powered Human–Machine Collaboration/Teaming' of the TRUSTY project [9]. Notwithstanding this, any piloting of the experimental set-up and methodology, before the eventual validation exercise will contribute to furthering our understanding of the HAIT technical design from an operational context.

It is appropriate to describe the activities of WP4 and WP5 here, as they represent an integral part of the project and the research, however their detailed planning and progress will not be captured in this ERP but elaborated on more in D3.2 'Report on the gap analysis including KPIs/KVIs and the development/ technical work plan' and in subsequent WP 4 and 5 deliverables.

Within WP4 the TRUSTY AI solution will be investigated and developed. This includes research and reporting on the following concepts:

a) **The use of Multimodal inputs and techniques in AI Automation:** In this task data analysis, MML techniques will be used to generate AI models;

b) **Robustness and resilience in ML models:** This task will develop ML models that are robust with high accuracy in prediction and decision-making;

c) **Transparent ML models incorporating interpretability, fairness and accountability:** This WP aims to comprise interpretable models for vision and image processing, especially for deep learning models;

d) **Human-centred explainable and active learning to the ML models:** In this task, a reflective sociotechnical approach will be used (1) to identify assumptions and requirements of the domain, (2) marginalize these requirements as components, (3) rationalize and realize the requirements, and (4) develop explainable ML models to embody the marginalized components.

Within WP5 the TRUSTY User-Interface for the AI will be investigated and developed. This includes research and reporting on the following concepts:

a) **Human Factors and measures relevant to HMI:** This task will identify the human factors of interest, relevant to enhancing the human-machine interaction effectiveness (i.e., HAIT). For each HF component of interest, the appropriate neurophysiological measure, feature to be examined, and related technologies to be used, will be identified;

b) **Human-In-the-Loop ML, UX for Human and Model Interaction**: This task will develop a passive Brain-Computer Interface, by using the mental states identified within the previous task. The output from the system will be used to trigger the AI model, to change specific modalities depending on the actual state of the operator;

c) **Framework for HAIT and AI-driven HMI**: This task will develop a framework to support improved HAIT within AI-driven human-computer interaction;

d) **Interactive Data Visualization and Multimodal HMI and GUI for Decision Support:** In this task the design and implementation of multimodal HMI and GUI for Decision Support will be used along with novel techniques for modelling and novel methods of interactive data visualization, including visual analytics techniques.

WP4 and WP5 will therefore produce the TRUSTY solution for validation and performance assessment, whilst exerting influence on how the experimental design and test mature.

The research approach in this study is exploratory and thus is looking to deepen the understanding of trust, and ultimately, acceptance concerning automated and AI assistance in the ATC environment. This is to better design for and manage these constructs during the inevitable increase in the use of AI in the aviation sector. Therefore, the research approach is one in which observations are made, in a scientifically rigorous experimental design, on the reaction of the users to the novel concept. The methodology used will attempt to elicit qualitative and quantitative data on the Human Performance KPA.

It should be noted that according to [6] at TRL 0 – TRL 2 performance data can be qualitative or quantitative. Some of the KPA cannot be measured quantitively at these low TRL levels and so subjective feedback will be generated on potential performance impacts.

The follow-on deliverable to this document will be the Exploratory Research Report (ERR), TRUSTY Deliverable 6.2, in which the results from this current research planning will be reported, in the form of validation analysis and an impact assessment. Also, in this follow-on deliverable there will be a proposed set of design guidelines for trustworthiness in AI, in the context of RDTs.

A review of the TRUSTY project's objectives shows that this ERP enables the following key project objectives:

Scientific, Innovation, And Research Objectives (SIO):

SIO#5: To provide a conceptual framework for building a trustworthy intelligent system. TRUSTY will undertake research and development concerning human-centric explainable-AI, incorporating fairness and accountability.

> *ERP Enabler: Conduct of validation exercise(s) and the subsequent production of D6.2. ERR, with impact analysis and guidelines.*

Impact and Societal Objectives (ISO):

ISO#1: To perform active engagement with main stakeholders in the definition of requirements and system conceptualization throughout the project. TRUSTY will incorporate AI research experts, RDTs operators, and ATM domain experts from different disciplines, and based on end users' needs, will co-create the HMI and the AI system using an integrated approach based on Human AI Teaming.

> *ERP Enabler: Conduct of validation exercise(s) and the subsequent production of D6.2. ERR, with impact analysis and guidelines.*

ISO#2: The TRUSTY project aspires to improve the transition process towards the use of RDTs by engendering more trust to RDTs operators, through transparent and explainable systems. Loss of depth perception and lack of auditory and tactile feedback may reduce RDT operators' situational awareness and impact on their skills developed, when compared with a conventional ATC tower environment. Increasing trust in novel technology, whilst complementing the experience and knowledge gained in the traditional setting, can provide a more safe and acceptable system.

> *ERP Enabler: Conduct of validation exercise(s) and the subsequent production of D6.2. ERR, with impact analysis and guidelines.*

The remaining 'Scientific, Innovation, And Research Objectives', 'Technological Objectives', 'User-centric Design Objectives', and 'Impact and Societal Objectives', detailed in the TRUSTY Grant Agreement [9] are met through other work packages, most substantially through WP4 and 5. Detailed here:

Table 4 Project Objectives met by other deliverables

Project Objective	*Details*	*Deliverable with provided evidence [9]*
SIO1	Provide a clear definition and a technical work plan of a trustworthy intelligent system with identified Key	D.3.1. Report on definition, specifications and SotA (M6);

	Performance Indicators (KPIs)/Key Value Indicators (KVIs) based on state of the art (SotA) study and gap analysis.	D 3.2 Report on the gap analysis including KPIs/KVIs and the development/technical work plan
SIO2	Provide a “Self-explainable” and “Self-learning system” for critical decision-making based on MML considering robust and resilient ML models to the tasks taxiway and runway inspection and misalignment warning.	D.4.1. Report on robustness and resilience MML with open-source models and database
SIO3	Provide ‘Transparent ML models’ incorporating interpretability, fairness, and accountability based on human centred XAI and active learning.	D.4.2. Report on the methodology of transparent ML models
SIO4	Provide an ‘Adaptive level of explanation’ regarding the user's cognitive state based on human factors and countermeasures relevant to the multimodal HMI and Human-In-the-Loop ML	D.5.2. Report on the methodology of human–machine teaming with human and Multimodal HMI and GUI with interactive data visualization
TO1	Robust and resilient MML models. An intelligent system for the decision-making tasks and monitoring of taxiways and runways in the RDTs domain adaptable ML models with high accuracy.	D.4.1. Report on robustness and resilience MML with open-source models and database
TO2	Transparent ML models with a human-centred explanation. A prototypical system of a proof-of-concept of the proposed intelligent system incorporating interpretability, fairness, and accountability.	D.4.2. Report on the methodology of transparent ML models
TO3	Framework for HAIT. A prototypical system of a framework incorporating human-AI-interaction (hAIi) and UX for human and ML model interaction.	D.5.2. Report on the methodology of HAIT through multimodal HMI and GUI with interactive data visualization
TO4	Smart HMI and GUI for intelligent decision support. A prototypical system of HMI and GUI will be developed incorporating interactive data visualization, data-driven storytelling, and a data exploration approach through visual analytics.	D.5.1. Report on the methodology of Human factors, countermeasures for HMI and UX for human and ML model Interaction

UCD1	Develop a trustworthy intelligent system, by using Human Centric AI explanation and HAIT that can be easily accepted by the RDTs operators	T4.4 Human-centred explainable and active learning to the ML models (reported on in D4.4 Report on the methodology of transparent ML models); T5.2 Human-In-the-Loop ML, UX for human and Model Interaction (reported on in D5.2 Report on the methodology of Human factors, countermeasures for HMI and UX for human and ML model Interaction)
ISO3	Communication, dissemination, and exploitation. Use the core project activities (research, Pilot case development, dissemination, standardization, and communication) to establish interactions and collaborations that will last beyond the project end.	D7.1 Dissemination, communication and exploitation strategy (DCE) plan and activities

3.3 Key R&I needs

The integration of Artificial intelligence and Machine Learning into safety-critical domains, such as aviation, still requires more attention and investigation in terms of technical robustness, safety, and human acceptance. Accordingly, the research and innovation needs associated with the expansion of the use of AI, specifically in the RDT environment are presented here.

R&I 1. Remote Digital Tower: The Remote Digital Tower concept is already in use in many parts of the world, providing the air traffic control operator with the capability of delivering a ATM service to remote airfields. RDTs can also be deployed for contingency purposes, for example when the in-situ tower is temporarily out of use (meteorological conditions with high wind, tower major renovation works, etc.). RDTs largely rely on video presentation and remote sensors to provide situational awareness, enabling the operator to fulfil their role.

> **Research Gap:** Due to concerns over safety and technological limitations, information monitoring and processing are mainly done today by the air traffic control operator. Such monitoring demands high levels of attention and vigilance but can generate boredom and fatigue in the operator. This is especially the case when a remote airfield may have very infrequent activities. To alleviate this situation, effectively designed Artificial intelligence for data processing could be envisaged.
>
> **Research Question:** How can AI technology assist the ATCO in providing a safe and effective ATM service to a remote airfield?

R&I 2. Trust in Artificial Intelligence: With the introduction of increasing complexity or autonomy in a system, trust in the technology by the user is called into question. “Trust is… a social construct that becomes relevant to human-machine relationships when the complexity of the technology defies our

ability to fully comprehend it". It can be seen as a 'willingness' to depend on the specific technology and is increasingly more important in situations in which negative safety consequences are possible.

> **Research Gap:** Due to their complexity and unpredictability, human-machine trust will impact significantly the use of AI-based tools. Moreover, it is not sufficient to exclusively deliver a reliable system with high levels of certainty in its output. In conjunction with this, carefully developed features are required for engendering trust and acceptance from a human perspective. The user interface design for AI can be challenging as some computations take a lot of time and thus specific incremental processing is required. It is also possible to add an abstraction layer to change the complex semantics of the algorithm parameters. This calls for a finely engineered solution where the human-machine interaction is optimised by using human factors to inform the design. This aspiration can pull from the current understanding on Human-AI Teaming to improve the communication pipeline between human and AI algorithms.
>
> **Research Question:** How is trust engendered through the design and integration of a 'trust-optimised' human-machine interface in an AI system? How are less human-compatible aspects of the AI technology overcome by the design of the user interface e.g. delays in outputs, lower certainty levels etc?

R&I 3. Adaptability of Transparency in the AI Output: Transparency in AI is of great interest for optimising trust and the capacity for AI to provide decision support capability. Allowing the user to have an accurate *perception* and a clear *comprehension* of the situation are the critical first and second steps for establishing good situational awareness, this is especially applicable and important in the use of AI.

Indeed, transparency in Machine Learning models is often coined 'interpretability' and 'explainability'. These two terms are often used interchangeably to depict the quality of a model to be interpreted and understood by the user. Interpretation usually refers to the whole inference process of the model and a single prediction by the model.

Moreover, using recordings from brain activity, it is possible to evaluate specific metrics that correlate with a variation of mental and emotional states of the user, such as workload, stress or vigilance, and this information can be used in real-time, to modify the behaviour of the AI and the presentation of information on the user-interface. This allows the AI ML model to adapt its behaviour, by considering the actual 'neurometrics' of the user.

> **Research Gap:** The appropriate level of transparency in an AI design requires further investigation as with increased transparency, more cognitive effort is required, by the user to efficiently analyse this additional information. Hence there is a balance between the amount and format of this extra information and the timely presentation of it, for example when the workload is already high, then the additional transparency is not useful. This calls for the level of transparency to be adapted according to user needs, in a particular situation and level of workload.
>
> **Research Question(s):** How should transparency be adapted in a dynamic design according to the level of workload and cognitive demands of the user? What factors represent the needs of the user in terms of transparency?

R&I 4. Involvement of Automation: Technology is critical in supporting the ATCO in their task in all ATM settings. Technology supports the reduction in the overall cognitive burden and complexity of the role and allows the operator to instead allocate precious cognitive resources to activities where the human component is essential. This technology can provide increasing levels of automation. Indeed,

AI can be classified according to the level of automation it provides [19]. In the case of TRUSTY, the proposed technology will provide automation at Level 1, namely i) support to information acquisition; ii) support to information analysis; and iii) decision-making support. This means that the AI is not necessarily making autonomous action in an ATM sense, but instead is collating data and providing information automatically.

> **Research Gap:** Historically, automation was aimed at taking work from the user in order to reduce workload, however it has since been realised that automation should optimise workload, not simply reduce it. This is because reducing workload can introduce problems such as boredom, complacency, and erosion of competence, and produce a considerable lack of situational awareness, especially in emergency situations. Maintaining stress and fatigue at optimal levels, and a workload level that allows the controller not to feel overburdened, but active and informed enough to respond appropriately, when necessary, requires a technological solution that is carefully designed to meet these needs. In this regard, having AI algorithms that are transparent, explainable, and reliable can have a strong impact on the successful introduction of automation.
>
> **Research Question:** How does AI *optimise* workload levels over time, whilst still reducing stress and fatigue and maintaining adequate situational awareness?

R&I 5. Multimodal Machine Learning: Humans perceive the world in a multimodal way, like seeing objects, hearing sounds, and feeling taste. To interpret the world like humans, AI can make progress and add benefits in interpreting multimodal information. This need brings the notion of multimodal machine learning, an engaging research area that integrates AI and multiple modalities. MML improves the capabilities of a ML model to analyse several data types with a resultant increase in accuracy.

> **Research Gap:** MML research began with audio-visual data but has now expanded to encompass text, signal, and sensor data. RDT's currently mostly rely on visual information from camera feeds however AI and MML can increase the modalities available to the ATCO by interpreting other aspects of the environment e.g. meteorological data.
>
> **Research Question(s):** What modalities can be used in the RDT environment to enhance AI output and support to the ATC operator? How are these modalities combined to give a single information feed?

R&I 6. Accuracy, Robustness and Resilience: To deliver an AI system that provides a trustworthy output, high accuracy and robustness must be key features of the AI design. To achieve the levels required, stability and constant high accuracy under different circumstances must be achieved. Specifically, AI model should be robust to small perturbations since real-world data contains diverse types of noise.

> **Research Gap:** Studies have shown that ML models can be fooled by small, designed perturbations, namely, adversarial perturbations. These perturbations are data inputs that cause a machine learning model to make a wrong prediction. They are imperceptible for humans, but sensitive enough for the model to change its prediction. Therefore, to minimize risk and remove bias, whilst providing resilience, the AI should encompass the capacity for adaptation in the ML models, to different situations and have an inherent ability to recover quickly. This should consider the effects of local constraints whilst mitigating the tendency for premature stopping (overfitting a model to data). To build safe and reliable ML models, investigating adversarial examples and the underlying reasons is urgent and essential for a trustworthy solution.

Research Question(s): What external factors will impact the AI ML in the ATM environment? How will these impact the accuracy of the ML model? How can the impact of these factors be mitigated?

In summary, there are several *research questions* that the TRUSTY project aims to explore, collected here from the discussion above on the Research and Innovation (R&I) needs:

Table 5 Project Research Questions

How can AI technology assist the ATCO in providing a safe and effective ATM service to a remote airfield?
How is trust engendered through the design and integration of a 'trust-optimised' human-machine interface in an AI system? How are less human-compatible aspects of the AI technology overcome by the design of the user interface e.g. delays in outputs, lower certainty levels etc?
How should transparency be adapted in a dynamic design according to the level of workload and cognitive demands of the user? What factors represent the needs of the user in terms of transparency?
How does AI *optimise* workload levels over time, whilst still reducing stress and fatigue and maintaining adequate situational awareness?
What modalities can be used in the RDT environment to enhance AI output and support to the ATC operator? How are these modalities combined to give a single information feed?
What external factors will impact the AI ML in the ATM environment? How will these impact the accuracy of the ML model? How can the impact of these factors be mitigated?

To answer these questions, the validation in this study will examine varying levels of Artificial Intelligence Certainty of detection and identification; and Human Machine Interface (Visualisation) Design (see section 4.3.1 for more details of these variables). Examining the design of the AI and the design of the HMI in different variants and recording the Human Performance and neurophysiological measures of relevant human factors (i.e. workload, stress, acceptability) outcome, it is envisaged that more insight will be provided against these R&D needs.

3.4 Estimated performance contributions

It is estimated that the TRUSTY project will contribute to several Key Performance Areas as defined in the DES performance framework [6], for the improvement of ATM. Those that are the focus of this validation exercise, are described as follows, in Table 6.

Table 6 Estimated performance contributions and summary of evidence

KPA	KPI	TRUSTY Impact	Evidence
SAFETY	SAF1 Total number of estimated accidents with ATM contribution	Medium	Qualitative
HUMAN PERFORMANCE	HP1 Consistency of human role with respect to human capabilities and limitations	High	Quantitative

	HP2 Suitability of the technical system in supporting the tasks of human actors	High	Quantitative
	HP3 Adequacy of team structure and team communication in supporting the human actors	High	Quantitative
	HP4 Feasibility with regards to HP-related transition factors	High	Quantitative

3.5 Initial and exit maturity levels

The initial TRL of the TRUSTY solution is zero as this is the initiation of the concept. It is envisaged that at the end of the project, the concept will have been progressed to level 1. This is captured in Table 7 below.

Table 7 Initial and exit maturity levels

Project/ Proposed SESAR solution(s) ID	Proposed SESAR solution(s) title	Initial maturity level	Exit maturity level	Reused validation material from past R&I Initiatives
SOL-TRUSTY	AI-Powered Situational Awareness for Remote Airfields (APSARA)	TRL0	TRL1	None

4 Exploratory research plan

4.1 Exploratory research plan approach

This section provides a description of how the exploratory research plan approach will progress the solution from the initial maturity level to the exit maturity level. To progress from TRL0 to TRL1, several components must be developed, as described in the DES SESAR Maturity Criteria and sub-criteria document [4] (criteria taken from [4] are in italics). These are discussed as follows:

a) *Formulation and documentation of the research hypothesis:* This will be explored and developed through the research and experimental planning for concept validation.

b) *Developing an innovative solution through the research activities and results:* Project work packages, namely WP4 and WP5 of the TRUSTY project will develop the TRUSTY HMI and AI algorithms, respectively. These work packages will progress the design of the technical solution from concept to early prototype, from which to conduct validation testing of the solution. The results of these early tests will produce results that aim to prove the concept as viable and beneficial to the ATM environment and will increase understanding within the areas of the R&I needs.

c) *Identifying and assessing the strengths and benefits of the solution, including also potential safety benefits:* This will be progressed through the validation exercise and by generating community interest through stakeholder engagement.

d) *Identifying and assessing the potential limitations, weaknesses and constraints of the solution under research, considering also potential safety and security considerations:* limitations, weaknesses and constraints will be continually considered through the project and validation exercise planning and conduct.

e) *Developing a contribution to strategic programme objectives e.g. performance ambitions identified at the ATM Master Plan level [1], strategic research and innovation agenda (SRIA) [2] and multiannual work programme (MAWP) [3]:* Contribution to higher level strategic objectives comes through the research and exploration of the application of novel technologies in support of Remote Digital towers specifically. The remote digital tower concept has been identified in strategic descriptions as a structure from which to enable the optimal use of air navigation service infrastructure and the use of scarce resources. As quoted in [1], “Virtual control centres and use of remote towers will allow a more efficient and flexible use of resources, substantially improving the cost efficiency of service provision”.

f) *Identifying, consulting and involving stakeholders in the assessment of the results. Documenting feedback in the project deliverables and generating interest in the proposed solution:* Stakeholder engagement was initiated in WP3 with stakeholder workshops and will be continued through the proposed validation exercise.

g) *Documenting recommendations for further scientific research:* Through documenting and reporting on lessons learned and by carefully exploring the performance of the solution, the project aims to have a clear idea of the future trajectory of the solution, by the end of the project, and reported in the ERR D6.2.

Much of the progress of the TRUSTY solution will be through WP4 and WP5 and will be measured in a final validation exercise. The validation exercise is aimed at proving the TRUSTY concept in a real-time simulated (RTS) environment involving (ATCO) human participants.

The focus of this exercise will be to evaluate the AI solution in the RDT setting, specifically looking at its performance as reflected in the levels of trust elicited in the user. This exercise aims to generate insight and progression against all R&I needs. An explanation as to how the exercise provides insight and progression against the R&I needs is described in Table 8. Independent bench testing will be conducted at the MDU site for the development of the AI, to address R&I needs 5 and 6, also detailed in Table 8. Both approaches reported for AI testing will provide probability measures of how accurate a prediction is under certain uncertainty conditions. Users will have good assumptions about the model performance, which will help them to make an informed decision.

Table 8 Explanation of how the validation exercise provides insight and progress in the identified R&I needs

R&I Need	Associated Research Question	Explanation of insight sought through the validation exercise
R&I 1. Remote Digital Tower	How can AI technology assist the ATCO in providing a safe and effective ATM service to a remote airfield?	The experimental set-up has been specifically designed to replicate the RDT dual operating mode set-up, as described in 2.2.1 and 5.1.1. Performance including human error and level of effectiveness will be measured/ observed in the scenarios, during the use of the AI tool.
R&I 2. Trust in Artificial Intelligence	How is trust engendered through the design and integration of a 'trust-optimised' human-machine interface in an AI system? How are less human-compatible aspects of the AI technology overcome by the design of the user interface e.g. delays in outputs, lower certainty levels etc?	The experimental design will test three different HMI visualisations aimed at varying levels of trust elicitation. These will be presented alongside varying levels of AI certainty to be able to observe levels of trust (and other human performance characteristics).
R&I 3. Adaptability of Transparency in the AI Output	How should transparency be adapted in a dynamic design according to the level of workload and cognitive demands of the user? What factors represent the needs of the user in terms of transparency?	The HMI visualizations are designed to represent three levels of transparency, namely; i) what is the AI showing me (can I tell if it's a bird on the runway or a bee in front of the camera)?; ii) Why is it showing me this (is it prioritising this attention-getter over others as it is more safety critical)?; iii) Do I need to know what it's showing me (would I judge the attention-getter as being critical to safety)? Alongside these varying 'treatments', workload and cognitive demands will be

R&I Need	Associated Research Question	Explanation of insight sought through the validation exercise
		measured to ascertain the impact of different visualisation designs (sections 4.3.1 and 5.1.6.2).
R&I 4. Involvement of Automation	How does AI *optimise* workload levels over time, whilst still reducing stress and fatigue and maintaining adequate situational awareness?	The AI tool in this study is designed around supporting the ATCO with ongoing monitoring of the RDT and decision-making in an event, thus it is aimed at the optimisation of workload and situational awareness across his/her tasks. Workload and situational awareness are key performance indicators being measured in this validation exercise.
R&I 5. Multimodal Machine Learning	What modalities can be used in the RDT environment to enhance AI output and support to the ATC operator? How are these modalities combined to give a single information feed?	In the RDT environment, different modalities, e.g., either video, audio, text or combinations of them, will be investigated to develop a multimodal machine learning system. Several aspects, especially Translation, Alignment and Co-learning, are of interest. Translation is understanding the relationship among the multimodal data; alignment identifies the relations between the scenarios and conditions related to events in the scenarios, and co-learning explores the advantages and limitations of each modality and uses that knowledge to improve the performances of models trained on a different modality. More detailed information will be added in D4.2. We will utilise the model's probability score to validate the accuracies of multimodal machine learning. The probability prediction score for true classes will be analysed to determine what would be the label of a data sample. This analysis will help determine the confidence of the model. The prediction probability distribution of any class of samples will first be selected. (Next, two threshold values will be defined by analysing the probability score. These threshold values are then used to create a condition that counts the number of samples that satisfy it. The percentage of samples

R&I Need	Associated Research Question	Explanation of insight sought through the validation exercise
		that satisfy the condition can be considered as the confidence of the model. We will also Bayesian approach for the validation which is a systematic method for updating beliefs based on new evidence. It calculates the posterior probability using the Bayesian theorem. The posterior probability for a sample is calculated for the true classes. Then, it is compared with the prediction probability score. The confidence of the model's prediction is then calculated by counting the similarity of each sample between prediction probability and posterior probability.
R&I 6. Accuracy, Robustness and Resilience	What external factors will impact the AI ML in the ATM environment? How will these impact the accuracy of the ML model? How can the impact of these factors be mitigated?	The robustness of the ML model will be measured using metrics such as accuracy, false positive rate, confusion matrix, etc. in terms of uncertainty quantification, i.e., to estimate how good or confident we are of the prediction produced by the ML model. In TRUSTY, we will consider the conformal prediction method to determine precise levels of confidence in predictions. The idea is to provide not a single prediction, but a prediction set with guaranteed coverage of the true class. Experiment setup. To validate the robustness of ML models, the conformal prediction will be evaluated in three steps: training, calibration, and prediction. 1. During the ML model's training, data will be split into training, testing, and calibration sets. The model will be trained on training data. 2. We will compute and sort the uncertainty scores, i.e., non-conformity scores for calibration data. For a new data instance, the non-conformity score measures how unusual the suggested prediction is

R&I Need	Associated Research Question	Explanation of insight sought through the validation exercise
		compared to the model output for other inputs. 3. A confidence level α = 0.1 will be used to achieve 90% coverage of true prediction in the conformal set. 4. The non-conformity score will be used to calibrate the prediction set for predicting outcomes of new data points, which will satisfy accuracy under certain uncertainty conditions (e.g., all outcomes are confident to have 90% probability coverage to fall under true classes).

Within this exercise, subjective opinion will be sought from the ATCO SMEs on the performance impact of the TRUSTY solution on all identified KPAs (Table 6). This data will be elicited in the debrief interview of the participant.

4.2 Stakeholders' expectations and involvement

The following table identifies relevant stakeholders, their involvement and why the scope of the research matters to them.

Table 9 Stakeholders' expectations and involvement

Stakeholder	Involvement	Why it matters to the stakeholder
Operational Establishments, ANSP's and Airline companies	Participation in validation exercises and interviews as operational Subject Matter Experts	The operators' needs are fundamental to the design and progression of novel technologies. Therefore, a human-centric approach is being adopted in this research, to optimise the design of future technologies. In this project the operator will have the opportunity to shape and influence this progress. This will ensure better, more effective integration of the technology into the operational environment.
Academic and Scientific Community	Advisory board participation and CDE audience.	Gain access to and ability to exploit the results of the research and details of experimental design.
Regulators and Policy Makers	Advisory board participation and CDE audience.	Insight into the RDT concept as a practical solution and progress towards integration of AI in the ATC/ATM environment. Insight into safety and design standards.

Industry	Advisory board participation and CDE audience.	Technical specification for future designs and more description of users' needs and requirements.

4.3 Validation objectives

This section presents a list of validation objectives which address and decompose the key R&I needs related to the project. More details on the pull-through from R&I needs can be found in Table 8. Human Performance validation objectives have been generated through use of the Human Performance Assessment Process.

Table 10 Validation Objectives, Success Criteria and Method of Measurement

KPA	KPI	VALIDATION OBJECTIVES	SUCCESS CRITERIA	Method of Measurement	Primary R&I Need
SAFETY	Safety 2 Suitability of technical system in supporting the tasks of human actors	OBJ-TRUSTY-SAF-ERP-02 The information provided is adequate for safely and effectively carrying out the task.	SC-TRUSTY-SAF-ERP-02-1 **System induced error** System tests show an adequate level of accuracy	System Tests in development	R&I 1 – Provision of a safe and effective ATM service. R&I 5. Multimodal Machine Learning. R&I 6 – Performance of the AI
			SC-TRUSTY-SAF-ERP-02-2 **System induced delay** System tests show an adequate level of timeliness	System Tests in development	R&I 1 – Provision of a safe and effective ATM service. R&I 5. Multimodal Machine Learning. R&I 6 – Performance of the AI
HUMAN PERFORMANCE	*HP1 The role of the human is consistent with human capabilities and limitations*				
		OBJ-TRUSTY-HP-ERP-01	SC-TRUSTY-HP-ERP-01-1 **Roles and Responsibilities**	Debrief questionnaire with Likert scale	R&I 1 – Provision of a safe and effective ATM service.

	01 Roles, Responsibilities and Procedures	ATCO role, responsibilities and procedures are acceptable and non-contradictory	The ATCOs judge the roles and responsibilities as clear, consistent and non-contradictory.		
		OBJ-TRUSTY-HP-ERP-01 ATCO role, responsibilities and procedures are acceptable and non-contradictory	SC-TRUSTY-HP-ERP-01-2 **Satisfaction** The ATCO judges job satisfaction as being acceptable.	Debrief questionnaire with Likert scale	R&I 1 – Provision of a safe and effective ATM service.
		OBJ-TRUSTY-HP-ERP-01 ATCO role, responsibilities and procedures are acceptable and non-contradictory	SC-TRUSTY-HP-ERP-01-3 **Procedures** The ATCOs judge the procedures and operating methods as being clear and non-contradictory.	Debrief questionnaire with Likert scale	R&I 1 – Provision of a safe and effective ATM service.
		OBJ-TRUSTY-HP-ERP-01 ATCO role, responsibilities and procedures are acceptable and non-contradictory	SC-TRUSTY-HP-ERP-01-4 **Task Allocation** Changes to task allocation within the human team are judged acceptable.	Debrief questionnaire with Likert scale	R&I 1 – Provision of a safe and effective ATM service.

	HP2 Suitability of the technical system in supporting the tasks of human actors				
	02 Human actors can achieve their tasks.	OBJ-TRUSTY-HP-ERP-02 The level of Human performance is acceptable.	SC-TRUSTY-HP-ERP-02-1 **Human Error** The ATCO is able to accurately fulfil the task.	• Hazard identification Analysis • SME Observational Grid / Rating Form • Video footage • Switch log data • Approach withdrawal (EEG)	R&I 1 – Provision of a safe and effective ATM service.
		OBJ-TRUSTY-HP-ERP-02 The level of Human performance is acceptable.	SC-TRUSTY-HP-ERP-02-2 **Task Performance** The ATCO is able to effectively fulfil their task.	• SME Observational Grid / Rating Form • Switch log data • Approach withdrawal (EEG) • Video footage	R&I 1 – Provision of a safe and effective ATM service. R&I 3 – AI Transparency
		OBJ-TRUSTY-HP-ERP-02 Level of Human performance is acceptable.	SC-TRUSTY-HP-ERP-02-3 **Timeliness** The ATCO is able to fulfil their task in a timely manner.	• Switch log data • SME Observational Grid / Rating Form	R&I 1 – Provision of a safe and effective ATM service. R&I 3 – AI Transparency
		OBJ-TRUSTY-HP-ERP-02 Level of Human performance is acceptable.	SC-TRUSTY-HP-ERP-02-4 **Workload** The ATCO judges workload in the task as being acceptable (according to cognitive/ physical	• WL scale • Attention/ Vigilance (EEG) • Mental Workload/ Effort (EEG)	R&I 4 – Optimisation of workload and situational awareness

			demands of the task, and task allocation).		
		OBJ-TRUSTY-HP-ERP-02 Level of Human performance is acceptable.	SC-TRUSTY-HP-ERP-02-5 **Trust** The ATCO judges trust in the new concept/ procedures, and the automated functions as being acceptable	• Pre/post Trust Subjective questionnaire • SME Observational Grid / Rating Form • SA probe with trust rating • Approach withdrawal (EEG)	R&I 2 – Trust in AI
		OBJ-TRUSTY-HP-ERP-02 Level of Human performance is acceptable.	SC-TRUSTY-HP-ERP-02-6 **Situational Awareness** The ATCO judges their situational awareness during the task as being acceptable	• Attention/ Vigilance (EEG) • Post condition Subjective SA Measure SART and/or; • Real-time Situation Awareness Probe and/or; • Observational SME SA Rating.	R&I 4 – Optimisation of workload and situational awareness

	03 The design of the human-machine interface supports the human in carrying out their tasks.	OBJ-TRUSTY-HP-ERP-03 The HMI Design is acceptable.	SC-TRUSTY-HP-ERP-03-1 **Usability** The ATCO judges the usability and information provision of the HMI as acceptable	• Approach withdrawal (EEG) • Subjective feedback – debrief questionnaire with Likert scale • UX Acceptance Index	R&I 2 – Trust in AI R&I 3 – AI Transparency
	HP3 Adequacy of team structure and team communication in supporting the human actors				
	04 Appropriate Human AI Teaming	OBJ-TRUSTY-HP-ERP-04 Level of Team performance is acceptable	SC-TRUSTY-HP-ERP-04-1 **Task Allocation – man/ machine** Changes to task allocation between the human and machine is judged acceptable according to task demands.	Debrief questionnaire with Likert scale	R&I 1 – Provision of a safe and effective ATM service.

		OBJ-TRUSTY-HP-ERP-04 Level of Team performance is acceptable.	SC-TRUSTY-HP-ERP-04-2 **Shared Mental Model** The ATCO is able to acquire an adequate mental model of the machine and its automated functions.	• Human Performance Envelope (Stress/WL/ Vigilance) (EEG/EDA) • Real-time Situation Awareness Probe	R&I 3 – AI Transparency R&I 4 – Optimisation of workload and situational awareness
		OBJ-TRUSTY-HP-ERP-04 Level of Team performance is acceptable.	SC-TRUSTY-HP-ERP-04-3 **Shared Situational Awareness** The ATCOs judge shared situational awareness during the task as being acceptable	• Post-hoc comparison of shared and complementary situation awareness for human-automation team.	R&I 4 – Optimisation of workload and situational awareness
		OBJ-TRUSTY-HP-ERP-04 Level of Team performance is acceptable.	SC-TRUSTY-HP-ERP-04-4 **Team Error** Errors observed in validation shows that the potential for team error is tolerable.	• Collective Human and Agent error	R&I 1 – Provision of a safe and effective ATM service.
		OBJ-TRUSTY-HP-ERP-04 Level of Team performance is acceptable.	SC-TRUSTY-HP-ERP-04-5 **Collective Timeliness** The Human/AI team are able to fulfil the team tasks in a timely manner.	• Collective Human and Agent timeliness	R&I 1 – Provision of a safe and effective ATM service.

	05 The communication between team members supports human performance.	OBJ-TRUSTY-HP-ERP-05 Team communications are acceptable.	SC-TRUSTY-HP-ERP-05-1 **Phraseology/communication content** ATCOs accept and judge the proposed phraseology/ communication content as being acceptable	Debrief questionnaire with Likert scale	R&I 1 – Provision of a safe and effective ATM service.
		OBJ-TRUSTY-HP-ERP-05 Team communications are acceptable.	SC-TRUSTY-HP-ERP-05-2 **Communication means and modalities** The ATCO judges that the communication means & modalities are acceptable	Debrief questionnaire with Likert scale	R&I 1 – Provision of a safe and effective ATM service.
		OBJ-TRUSTY-HP-ERP-05 Team communications are acceptable.	SC-TRUSTY-HP-ERP-05-3 **Communication Load** ATCOs accept and judge the communication load as being acceptable in normal, abnormal and degraded mode operations.	• Debrief questionnaire with Likert scale • Vigilance/Attention (EEG)	R&I 4 – Optimisation of workload and situational awareness

4.3.1 Dependent, Independent and Control Variables

The ***independent*** variables in the TRUSTY project are:

IV1 – Artificial Intelligence Certainty of detection and identification: two levels of certainty e.g. about 60% certainty and about 90% certainty.

These two values are designed to represent a highly accurate system e.g. 90%/100%, in which the user has reason to trust the system, and a lesser accurate system e.g. 50%/60%, which will mimic a system in which trust is not justified. These values will represent two different situations to see how the participants respond in terms of trust.

IV2 – Human Machine Interface (Visualisation) Design: based on Transparency, Explainability and a Shared Mental Model.

These three aspects will allow greater insight into how the visualisations encourage or discourage trust, the following gives an illustration of the three factors:

> *Transparency*: What is the AI showing me (can I tell if it's a bird on the runway or a bee in front of the camera)?
>
> *Explainability*: Why is it showing me this (is it prioritising this attention-getter over others as it is more safety-critical)?
>
> *Shared mental model:* Do I need to know what the AI is showing me (would I judge the attention-getter as being critical to safety)?

These two variables will be fixed, in 2 and 3 configurations respectively, to measure their impact through the dependent variables listed below.

The ***dependent*** variables are presented in Table 11. These are the variables being measured and/or observed and are those that change depending on alterations in the independent variable. These variables have been identified as factors that give insight into the impact of the independent variables selected and that provide insight according to the R&I needs. Trust is a key variable in this project however trust is a multi-faceted factor and so other Human Performance measures need to be taken and correlated with the trust measurement to understand what aspects influence trust. Moreover, the nature of the dual-task experimental set up is novel and so other human performance measures are necessary to fully articulate the baseline scenario.

In Table 11 the associated validation criteria number, from Table 10, has been listed for easy reference to the validation objective, success criteria, method of measurement and relevant R&I need.

Table 11 Dependent Variable and associated Validation criteria reference number

Dependent Variable	*Validation criteria reference number (see* Table 10*)*
Task Performance	HP-ERP-02-2
Timeliness	SAF-ERP-02-2; HP-ERP-02-3; HP-ERP-04-5
Workload	HP-ERP-02-4
Trust	HP-ERP-02-5

Situational Awareness (individual and shared)	HP-ERP-02-6; HP-ERP-04-3
Acceptability	SAF-ERP-01-1; HP-ERP-01-1; HP-ERP-01-2; HP-ERP-01-3; HP-ERP-01-4; HP-ERP-04-1; HP-ERP-05-1; HP-ERP-05-2; HP-ERP-05-3
Mental model formation	HP-ERP-04-2
Human Error	SAF-ERP-02-1; HP-ERP-02-1; HP-ERP-04-4
Usability	HP-ERP-03-1;

4.4 Validation assumptions

This section provides validation assumptions that are applicable to this ERP and validation exercise, and which may have an impact on the validation results.

Table 12 Validation assumptions overview

ID	Assumption title [5]	Assumption description
AssumID#1	Traffic Characteristics	It is assumed that the traffic in the RDT is minimal and requires only a supervisory air traffic management service
AssumID#2	Airport Characteristics	It is assumed that the ATCO has access to a secure, real-time video link at the conventional tower, from the RDT.
AssumID#3	Ground Tools/ Technology	It is assumed that the RDT has video capability for runway/ taxiway monitoring. It is assumed that the conventional ATC tower control room has the capacity for monitoring in-situ air traffic and remote air traffic.
AssumID#4	Procedures in Place	It is assumed that there are normal, abnormal and emergency operating procedures in place to support the technology. It is assumed that roles and responsibilities are defined, and tasks allocated appropriately.
AssumID#5	Human performance	It is assumed that there is a training structure in place that is capable of training on AI technologies.
AssumID#6	Regulatory	It is assumed that regulation, legislation and certification is mature enough to manage the integration of AI technologies.

4.5 Validation exercises list

The following details the validation exercise required to successfully achieve the exit maturity level.

Table 13 Validation Exercise TVAL.10.1-TRUSTY-0434-TRL1 Layout

Identifier	TVAL.10.1-TRUSTY-0434-TRL1
Title	Real Time Simulator Testing of the TRUSTY concept with SME ATCOs
Description	TRUSTY concept AI and MML technology for the detection, processing and notification of airfield events with increased risk to the pilot.

	Proof of the TRUSTY concept in a real-time simulated environment involving (ATCO) human-participants. The objective of this exercise will be to evaluate the AI solution in a simulated RDT setting, focussing on its impact on operational safety and effectiveness in normal and abnormal conditions. Human performance analysis will be conducted, focussing considerably on the aspect of trust.
KPA/TA addressed	All
Addressed expected performance contribution(s)	Increased situational awareness, optimisation of workload whilst, improved timeliness of task execution, acceptance of human-AI task allocation, increased trust and tolerable levels of human error.
Maturity level	TRL1
Use cases	See section 5.1.4 Validation scenarios
Validation technique	Use of professional and student ATCOs for SME feedback in a representative environment
Validation platform	Aeronautical Computer Human Interaction Laboratory (ACHIL)
Validation location	ENAC, France
Start date	January 2025
End date	March 2025
Validation coordinator	ENAC, Deep Blue
Status	Under preparation
Dependencies	Availability of ATCO participants.

4.6 Validation exercises planning

Validation Exercises' Schedules

Title	Identifier	Start Date	End Date
Exercise A: Real Time Simulator Testing of the TRUSTY concept with SME ATCOs	TVAL.10.1-TRUSTY-0434-TRL1	January 2025	March 2025

4.7 Deviations with respect to the SESAR 3 JU project handbook

There are no known events and decisions that have led to a deviation with respect to the SESAR 3 JU project handbook.

5 Validation exercises

5.1 Validation Plan Exercise A: Real Time Simulator Testing

5.1.1 Validation Exercise Description and Scope

This validation exercise addresses the concept of the TRUSTY solution in a real time validation exercise, through the use of a representative flight simulator facility and employment of operational ATCOs in the test of the technology in various scenarios.

The operational context of this exercise is rooted in the concept of delivering an air traffic service through a 'multiple operating mode'. This involves provision of the service to multiple airports simultaneously from a single digital tower facility. To explore this concept, a theoretical approach involving a dual-tasks experimental set up has been designed. The simulated setting will require the ATCO participant to manage air traffic in-situ at a conventional, co-located tower, alongside supervising events occurring at a remote airfield.

Over a series of runs, the ATCO will be exposed to differing conditions in which the independent variables of 'AI certainty of detection', and 'HMI visualisation design' will be modified to measure the resultant dependant variables listed in section 4.3.1.

Each run will involve the presentation of a number of 'events. These events will be presented at the in-situ airfield, which will represent standard operations, and at the remote airfield which will be less benign in nature. Examples of events at the remote airfield include bird hazards; vehicle, person or animal on the runway; FOD; Engine fire; electrical failure; or wind shear.

The AI and accompanying visualization will be presented on the screens of the remote airfield to ascertain the performance of the TRUSTY technology in assisting the ATCO in handling the events. The AI will be taking multi-modal data from visual and auditory feeds for anomaly detection, namely video footage and pilot-ATC communications, at the remote airfield. The visualisations will be of varying designs exhibiting different levels of transparency and explainability of the output data from the AI.

In this exercise, human performance is being measured across the numerous runs. As the dual-task, multiple operating mode paradigm is novel, the experimental design will include a baseline condition in which there is no AI/visualisation component. Thus, the level of human performance in this setting, including workload and situational awareness will be measured, including subjective feedback on this paradigm as an operational concept. During the presentation of the AI/visualisation, human performance measurements will be focused on trust, acceptance and suitability of the technology.

The platform being used for the simulation is the Aeronautical Computer-Human Interaction Laboratory (ACHIL) situated at ENAC Toulouse. This facility consolidates a variety of simulators, representing numerous ATM positions in a single location, including en-route, approach, and tower positions for ATC, a cockpit simulator, and a supervision room. Details of the facility are described in section 5.1.7.

This exercise will progress the TRL level of the TRUSTY concept by contributing to the strategy described in 4.1, specifically item numbers c) Identifying and assessing the strengths and benefits of the solution; and d) Identifying and assessing the potential limitations, weaknesses and constraints of the solution.

5.1.2 Stakeholder's expectations and benefit mechanisms addressed by the exercise

Stakeholders' expectations, involvement and benefits are described in Table 9.

5.1.3 Validation objectives

The validation objectives of this exercise are those related to the Human Performance KPA and are detailed below in Table 14.

Table 14 Validation objectives addressed in the validation exercise

#	SESAR solution validation objective (same as Exercise Validation Objective)	SESAR solution success criteria (Same as Exercise Success Criteria)	Coverage of SESAR solution validation objective in exercise A
1.	OBJ-TRUSTY-HP-ERP-01 The ATCO's role, responsibilities and procedures are acceptable and non-contradictory	SC-TRUSTY-HP-ERP-01-1 **Roles and Responsibilities** The ATCOs judge the roles and responsibilities as clear, consistent and non-contradictory.	Full Coverage
2.	OBJ-TRUSTY-HP-ERP-01 The ATCO's role, responsibilities and procedures are acceptable and non-contradictory	SC-TRUSTY-HP-ERP-01-2 **Satisfaction** The ATCO judges job satisfaction as being acceptable.	Full Coverage
3.	OBJ-TRUSTY-HP-ERP-01 The ATCO's role, responsibilities and procedures are acceptable and non-contradictory	SC-TRUSTY-HP-ERP-01-3 **Procedures** The ATCOs judge the procedures and operating methods as being clear and non-contradictory.	Full Coverage
4.	OBJ-TRUSTY-HP-ERP-01 The ATCO's role, responsibilities and procedures are acceptable and non-contradictory	SC-TRUSTY-HP-ERP-01-4 **Task Allocation** Changes to task allocation within the human team are judged acceptable.	Full Coverage
5.	OBJ-TRUSTY-HP-ERP-02 Level of Human performance is acceptable.	SC-TRUSTY-HP-ERP-02-1 **Human Error** The ATCO is able to accurately fulfil their task.	Full Coverage

6.	OBJ-TRUSTY-HP-ERP-02 Level of Human performance is acceptable.	SC-TRUSTY-HP-ERP-02-2 **Task Performance** The ATCO is able to effectively fulfil their task.	Full Coverage
7.	OBJ-TRUSTY-HP-ERP-02 Level of Human performance is acceptable.	SC-TRUSTY-HP-ERP-02-3 **Timeliness** The ATCO is able to fulfil their task in a timely manner.	Full Coverage
8.	OBJ-TRUSTY-HP-ERP-02 Level of Human performance is acceptable.	SC-TRUSTY-HP-ERP-02-4 **Workload** The ATCO judge's workload in the task as being acceptable (according to cognitive/ physical demands of the task, and task allocation).	Full Coverage
9.	OBJ-TRUSTY-HP-ERP-02 Level of Human performance is acceptable.	SC-TRUSTY-HP-ERP-02-5 **Trust** The ATCO judges trust in the new concept/ procedures, and the automated functions as being acceptable	Full Coverage
10.	OBJ-TRUSTY-HP-ERP-02 Level of Human performance is acceptable.	SC-TRUSTY-HP-ERP-02-6 **Situational Awareness** The ATCO judges their situational awareness during the task as being acceptable	Full Coverage
11.	OBJ-TRUSTY-HP-ERP-03 The HMI Design is acceptable.	SC-TRUSTY-HP-ERP-03-1 **Usability** The ATCO judges the usability and information provision of the HMI as acceptable	Full Coverage
12.	OBJ-TRUSTY-HP-ERP-04 Level of Team performance is acceptable	SC-TRUSTY-HP-ERP-04-1 **Task Allocation – man/ machine** Changes to task allocation between human and machine is judged acceptable according to task demands.	Full Coverage

13.	OBJ-TRUSTY-HP-ERP-04 Level of Team performance is acceptable.	SC-TRUSTY-HP-ERP-04-2 **Shared Mental Model** The ATCO is able to acquire an adequate mental model of the machine and its automated functions.	Full Coverage
14.	OBJ-TRUSTY-HP-ERP-04 Level of Team performance is acceptable.	SC-TRUSTY-HP-ERP-04-3 **Shared Situational Awareness** The ATCOs judge shared situational awareness during the task as being acceptable	Full Coverage
15.	OBJ-TRUSTY-HP-ERP-04 Level of Team performance is acceptable.	SC-TRUSTY-HP-ERP-04-4 **Team Error** Errors observed in validation shows that the potential for team error is tolerable.	Full Coverage
16.	OBJ-TRUSTY-HP-ERP-04 Level of Team performance is acceptable.	SC-TRUSTY-HP-ERP-04-5 **Collective Timeliness** The Human/AI team are able to fulfil the team tasks in a timely manner.	Full Coverage
17.	OBJ-TRUSTY-HP-ERP-05 Team communications are acceptable.	SC-TRUSTY-HP-ERP-05-1 **Phraseology/communication content** ATCOs accept and judge the proposed phraseology/ communication content as being acceptable	Full Coverage
18.	OBJ-TRUSTY-HP-ERP-05 Team communications are acceptable.	SC-TRUSTY-HP-ERP-05-2 **Communication means and modalities** The ATCO judges that the communication means & modalities are acceptable	Full Coverage

19.	OBJ-TRUSTY-HP-ERP-05 Team communications are acceptable.	SC-TRUSTY-HP-ERP-05-3 **Communication Load** ATCOs accept and judge the communication load as being acceptable in normal, abnormal and degraded mode operations.	Full Coverage

5.1.4 Validation scenarios

The validation scenarios for the TRUSTY project are designed to be consistent with the operational concept outlined, focusing on the dual-task scenario involving ATM provision at an in-situ airfield and airfield supervision at a remote airfield. The validation exercise includes both a reference scenario (baseline) without the TRUSTY solution, and a scenario with the TRUSTY solution integrated into the RDT displays. These are described as follows:

5.1.4.1 Reference scenario(s)

***Description**:* The reference scenario represents a baseline operational environment where the ATCO performs tasks without the assistance of XAI in the RDT. Scenario specifics are:

Airport Specifics: The simulated environment is based on Toulouse-Blagnac Airport for ATM tasks and RDT Murhet-Lherm Aerodrome for airfield monitoring.

Airspace Specifics: Standard airspace layout and regulations are applied without any AI-driven enhancements.

Traffic Specifics: The traffic scenario includes regular aircraft movements requiring ATCOs to manage take-offs, landings, and taxiway usage.

Operational Setup:

The RTS Location is the ACHIL platform at the ENAC facilities, Toulouse.

The ATCO will conduct a primary task, involving a conventional tower simulation ATCO responsibilities include: • Authorizing aircraft for take-offs and landings. • Directing taxiway usage to prevent congestion and ensure smooth transitions. • Coordinating with ground services for efficient aircraft movement.
The ATCO will conduct also a secondary task, involving the RDT Murhet-Lherm airfield ATCO responsibilities include: • Monitoring airfield events such as unauthorized runway incursions and anomalies. • Managing runway operations manually based on visual and auditory cues. • Responding to alerts without XAI assistance, relying solely on human judgment and existing monitoring systems.

Objective:

This scenario aims to establish a baseline for ATCO performance and trust levels in an environment without XAI support, focusing on the traditional methods of air traffic and airfield management, whilst still involving the RDT.

5.1.4.2 Solution scenario(s)

***Description**:* The solution scenario introduces an XAI system to the RDT, enhancing the operational environment with advanced monitoring and decision-support capabilities. Scenario specifics are:

Airport Specifics: Same as the reference scenario, based on Toulouse-Blagnac Airport for ATM tasks and RDT Murhet-Lherm Aerodrome for airfield monitoring.

Airspace Specifics: Standard airspace layout with AI-driven enhancements for improved situational awareness.

Traffic Specifics: Identical to the reference scenario, maintaining consistency for comparison purposes.

Operational Setup:

The RTS Location is the ACHIL platform at the ENAC facilities, Toulouse.

The ATCO will conduct a primary task, involving a conventional tower simulation

ATCO responsibilities include (as a repeat of the reference scenario):

- Authorizing aircraft for take-offs and landings.
- Directing taxiway usage to prevent congestion and ensure smooth transitions.
- Coordinating with ground services for efficient aircraft movement.

The ATCO will conduct also a secondary task, involving the RDT Murhet-Lherm airfield with XAI Integrated in the ATC system

XAI Capabilities include:

Object Detection: XAI identifies and tracks common and anomalous objects (e.g., birds, vehicles) on the runway.

Event Detection: XAI system detects significant events and anomalies, providing real-time alerts to the ATCO.

Alert Mechanisms: XAI generates auditory and visual alerts to capture the ATCO's attention to critical situations.

Data Analysis: XAI analyses video, audio, and weather data to identify high-risk situations and supports decision-making.

ATCO responsibilities include:

- Interacting with the XAI system to manage runway operations based on real-time data and AI-driven insights.
- Responding to AI-generated alerts, ensuring prompt and efficient handling of critical situations.

• Using AI-provided data to enhance situational awareness and make informed decisions aligned with safety protocols.
Objective: The solution scenario aims to demonstrate the impact of XAI integration on ATCO performance and trust levels. By comparing this scenario to the reference scenario, the study will assess how XAI enhancements improve situational awareness, operational efficiency, and trust in XAI systems.

5.1.5 Exercise validation assumptions

Validation assumptions are the same as those recorded in Table 12.

5.1.6 Limitations and impact on the level of significance

5.1.6.1 Limitations and Applicability of the Experimental Results

The results of the research are applicable to safety critical, highly regulated environments in which complex traffic situations, with multiple agents are present. This could be other areas of ATM, aviation in general, ground transportation, congested maritime environments and drone operations. The population of users that the results can be transferred to are those who are required to gain complete and timely situational awareness in high workload environments, and who are required to utilize technologically advanced decision support and automation tools, as an integrated part of their operational equipment.

However, one limitation to the specific application of the research results to the ATM sector, is that of the use of the dual-task scenario described. This operating mode, in which the same ATCO provides in-situ ATC and remote airfield supervision is a theoretical construct and currently does not exist in the ATM environment. At present, two different ATCOs would provide this service, one for each airfield whether in-situ or remotely located. Therefore, it is not ecologically representative, however it has been envisaged as a probable future scenario, according to strategic vision [1][18].

A further limitation is that the AI algorithm will not be integrated into the simulated environment as it is not possible to integrate the AI into the simulated platform. Therefore, the testing of the HMI and the development of the AI are being done somewhat separately. Instead, representative outputs of the AI will be simulated in the scenarios through the RTS platform.

5.1.6.2 Managing Potential Sequence Effects

Testing a novel operating mode will mean that the setting and role will be new to the ATCO participants and so there will possibly be a learning effect from this operating mode, but also from the use of the XAI. To mitigate the effects of this, a familiarisation period will be included in the scheduling. Moreover, a baseline scenario is included in the testing, which is without the TRUSTY solution, to ascertain the level of human performance being achieved in this novel operating mode. This then allows for a reference point to correlate any effects from the addition of the XAI solution.

As there are two independent variables with three conditions, a Latin Square matrix has been designed to assess every combination of conditions. To prevent a sequence effect in the occurrence of events in the experimental runs, randomization will occur for the presentation of events to the participant.

The Latin square for the experimental design is presented in Figure 1. In this figure, AI Technical Reliability is a level of certainty in detection, afforded by the AI. This factor is aimed at eliciting a response in the user on the *trustworthiness* of the AI from an *accuracy* perspective.

As a key for this variable:

Reliability 90%	9 out of 10 items detected
Reliability 60%	6 out of 10 items detected
No AI	Standard ATCO Monitoring

Also, in Figure 1 the accompanying independent variable is the type of information presentation or 'visualization'. This factor is aimed at eliciting a response from the user, in terms of the following

Transparency: What is the AI showing me (can I tell if it's a bird on the runway or a bee in front of the camera)?

Explainability: Why is it showing me this (is it prioritising this attention-getter over others as it is more safety critical)?

Shared mental model: Do I need to know what the AI is showing me (would I judge the attention-getter as being critical to safety)?

Figure 1 Latin square for the experimental design

		AI Technical Reliability Treatment		
		90%	60%	No AI
Visualization Treatment	All objects detected are shown	1) All/90	2) All/60	7) No
	All objects are shown with indication of what the object is	3) All+label/90	4) All+label/60	
	Only objects that are likely to have a safety implication are shown with explanation of why it's being shown	5)Select+exp/90	6) Select+exp/60	

In terms of the function of the system in these conditions, examples of the preliminary design of the visualisations, representing the output of the AI, are as follows (however the design is due to be finalised in D5.4 Report on the methodology of human– machine teaming with human and Multimodal HMI and GUI with interactive data visualization due in M18):

Condition 1) All/ 90% - Represents the AI showing the ATCO all the objects which have been detected. Differentiation of objects can be shown through colour. In this example the AI reliability is high.

Figure 2 Condition 1)

Condition 3) All+label/90% - Represents a level of explainability, showing all the objects detected but also a label is present to indicate to the ATCO what the item is (relevant in cases when it is not obvious) and the level of certainty of classification of that object. In this example the AI reliability is high.

Figure 3 Condition 3)

Condition 5) Select+label/90% - Represents the objects detected with some explanation, but also this visualisation is designed to meet the expectations and mental model of the ATCO who has a need for safety critical objects to be detected, but not necessarily all objects. In this example the AI reliability is high.

Figure 4 Condition 5)

Condition 7) No AI – represents the reference scenario in which the AI is absent as well as the visual representation.

Figure 5 Condition 7)

Condition 2) All/60 - Represents the AI showing the ATCO all the objects that have been detected, but as the AI reliability is lower in this condition, then irrelevant objects are detected and also important ones have been missed e.g. the birds.

Figure 6 Condition 2)

Condition 4) All+label/60 - Represents the AI showing the ATCO all objects detected, with added explanation from the label, but as the AI reliability is lower in this condition, then irrelevant objects have been detected, and the certainty of the classification is low (and incorrect in this case) and also important objects have been missed e.g. the birds.

Figure 7 Condition 4)

Condition 6) Select+label/60% - Represents the objects which are considered safety critical are displayed with some explanation, designed to meet the expectations of the ATCO, but as the AI reliability is lower in this condition, then non safety critical objects have been detected and incorrectly labelled and also important objects have been missed e.g. the birds.

Figure 8 Condition 6)

5.1.6.3 Participants

To run through every combination of the variables, a test matrix requires 7 runs, detailed in Table 15. A participant pool of five has been chosen as most data being collected is subjective and there are diminishing returns in terms of insight gained, beyond data collected from five participants [22]. In terms of participant recruitment and required demographic, this information is described in detail in the TRUSTY project deliverable D2.2 H - Requirement No. 1. [23].

Table 15 Participant Test Schedule

ATCO	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Run 7
1	All/90	All+label/ 60	Select+ exp/90	No	All/60	All+label/ 90	Select+ exp/60
2	Select+ exp/60	All/90	All+label/ 60	Select+ exp/90	No	All/60	All+label/ 90
3	All+label/ 90	Select+ exp/60	All/90	All+label/ 60	Select+ exp/90	No	All/60
4	All/60	All+label/ 90	Select+ exp/60	All/90	All+label/ 60	Select+ exp/90	No
5	No	All/60	All+label/ 90	Select+ exp/60	All/90	All+label/ 60	Select+ exp/90

Key:

All	All objects detected are shown
All + label	All objects are shown with indication of what the object is
Select + Exp	Only objects that are likely to have a safety implication are shown with explanation of why it's being shown
No	No AI (baseline)

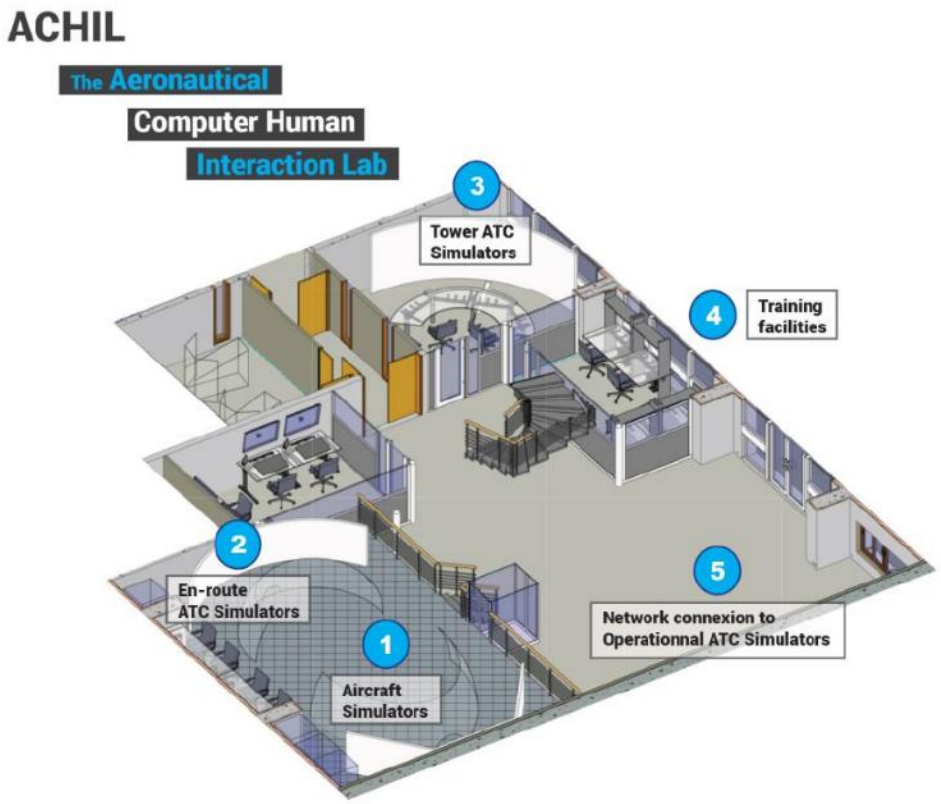
60	60% AI Reliability
90	90% AI Reliability

5.1.7 Validation exercise platform / tool and validation technique

5.1.7.1 ACHIL Platform

The Aeronautical Computer-Human Interaction Laboratory (ACHIL) platform Figure 9, housed at the ENAC facilities, consolidates a variety of simulators for numerous ATM positions in a single location: en-route, approach, and tower positions for ATC, cockpit simulators, and a supervision room (please refers to AEON D4.1 for more information see [24] The close integration of air and ground positions enables a concentrated focus on air-ground collaboration. The simulation tools used facilitate the creation of a highly realistic environment for operational experts (including AMAN, TCAS, Safety Nets, Aircraft models, etc.). Leveraging a specialized middleware, the system's flexibility also accommodates research needs and rapid prototyping, providing the capability to swiftly establish a complete operational environment for exploring new ideas and concepts.

Figure 9 ACHIL simulation facilities



In the frame of the TRUSTY project, the main position to be used will be the ground tower position (Figure 10) from the control tower of Toulouse Blagnac airport (LFBO) (Figure 11). The tower position uses RealTwr (RealTower) (https://realtwr.fr/) for the out of the window view.

Figure 10 Ground tower position with Real Tower view

Figure 11 Tower position with Real Tower view from Toulouse Blagnac airport (LFBO)

Overall, the software developed for the simulation platform covers both its construction and operation. Dedicated tools have been created to generate input data for simulations, while others are designed for the simulations themselves. The platform is adaptable to various airports, with data provisioning developed to be sufficiently generic. The primary source for airport maps, including background images and routing networks with taxi-lanes, taxiways, and run-ways, comes from open-source data for the X-Plane flight simulator. The AEON team [24] developed a Python program to extract information from these files and generate an SVG file with all the necessary details. External sources like Open Topo Data (https://www.opentopodata.org/) were used to obtain intersection altitudes and compute taxiway slopes. The SVG file structure defines edges as portions of taxiways or runways between intersections, with additional information such as length, slope, and turn radius. A visual editor is being developed to facilitate the definition of traffic rules. The IVY protocol (https://www.eei.cena.fr/products/ivy/), a simple set of libraries and programs for text message communication between applications, is utilized. It is open-source and supports multiple programming languages on various operating systems. IVY is effective even with many agents and is used in numerous research projects, including those in air traffic control and human-computer interaction.

The simulation engine, an IVY agent, sends messages for radar tracks with position information every second and responds to flight plan data requests. Scenarios can be created based on flight schedules,

and the engine generates corresponding data in the radar tracks database. For inbound flights, radar tracks are computed from 10 miles out until the aircraft vacates the runway, after which a pseudo pilot coordinates taxiing with ATCO. For departing flights, radar tracks are generated at the parking position, with the pseudo pilot managing pushback and taxiing orders. The simulation stores flight plans, including departure and arrival airports, aircraft type, equipment, runway, parking position, and timing information. It also maintains the status of each flight and updates agents on changes.

Radio communications utilize an IVY agent implementing the AudioLAN protocol, adding live voice modifications to simulate different voices for aircraft, despite only a few pseudo pilots being in charge. For evaluation, an additional IVY agent will handle data logging, with every piece of information timestamped in real-time. This architecture allows for comprehensive analysis, with relevant agents sending log messages in real-time during simulation runs.

5.1.7.2 Remote Digital Tower Platform (Murhet-Lherm Aerodrom)

The ACHIL platform has an extension on Muret airfield (LFBR). There a mast can host different sensors; it currently holds 4 cameras and a 'light detection and ranging' (LIDAR) Figure 12.

Figure 12 Cameras and LIDAR from the Muret airfield

The data are multiplexed and live streamed to ENAC Toulouse premises where the videos are stitched and processed by a neural network to detect vehicles and persons on the apron.

Figure 13 RealTwr data streams from video feed of Muret Airfield

In the Remote Tower room Figure 13, the video is displayed on a 180°, round screen and 2 simulated control positions are fed with the computed radar data. The next step is to install and connect the same system in Carcassonne airport (LFMK).

5.1.8 Data collection and analysis

5.1.8.1 Human Performance Data

The results from the validation activities will incorporate a combination of quantitative and qualitative data collection methods, ensuring a comprehensive evaluation of the validation objectives. Relevant test metrics are detailed in the table of validation objectives Table 10. Quantitative data will be gathered primarily through structured questionnaires utilizing Likert Scales (either 5-point or 7-point, depending on the original versions of the validated instruments). These scales will assess constructs such as Workload, Situational Awareness and Usability. Additionally, Trust will be evaluated as described in the TRUSTY project deliverable D3.2 [16] following the work performed by [25] in their meta-analysis. Therefore, the trust measurement will encompass different dimensions, namely situational, dispositional and shared mental model, which will be assessed through validated questionnaires already adopted for similar scopes. The numerical data derived from these scales will facilitate precise measurement and comparison of data.

Human performance, including human error, will be observed and rated by using observational methods, video footage, switch logs and debriefing questions. Instantaneous Self-Assessment (ISA) measurements of Human Performance including situational awareness and trust will be made, when possible, whilst maintaining minimising disruption to the ecological nature of the scenario.

In addition to the quantitative approach, qualitative data will be collected through debriefing sessions. These sessions will provide in-depth insights into participant experiences and perceptions, offering valuable context and depth to the quantitative findings. This combination of data types will enable a holistic understanding of the validation results, capturing both measurable outcomes and experiential nuances.

The use of validated questionnaires ensures the reliability and validity of the data collected, providing standardized measures of key constructs. This approach is particularly suitable for achieving the validation objectives as it allows for the assessment of specific metrics critical to evaluating trust, workload, situational awareness, and usability. The qualitative debriefing sessions will capture the nuances of participant experiences, informing improvements and refinements in the project's implementation. Eventual additional ad-hoc questionnaires for emerging testing needs, in the months prior to the beginning of the validation exercises, may be adopted.

The TRUSTY project is designed with a single validation activity, streamlining the validation process to provide a clear, focused assessment of the project's outcomes without multiple iterations. The project requires specific data types, including video stream input for remote Tower control and self-report measures from validated and ad-hoc questionnaires. The video stream will be provided by the partner ENAC, whilst the self-report measures will be gathered by the Test & Validation Lead Deep Blue, through questionnaire administration.

All collected data will be processed and cleaned. This includes handling missing data, normalizing scores where appropriate, and preparing the data for statistical analysis to ensure the accuracy and integrity of the results. The output of the experiments will be recorded in various formats, including system logs to capture detailed performance data, questionnaires documenting participant responses, and audio or video recordings of debriefing sessions supplemented by written notes.

Objective data from neurometric sensors (see 5.1.8.2) will endow additional and objective information about ATCO’s mental states while dealing with the validation scenarios. The temporal resolution of the mental states' indicators (mental workload, stress, vigilance and approach-withdrawal) will be adjusted according to other measurements so that correlation analyses will be performed with other kind of measures (e.g. subjective, behavioural).

Further quantitative data will be obtained from system data recorded during each session. These data contain information on the simulation platform. These data will be used to analyse the performance of the system and the input of the participant.

5.1.8.2 Neurophysiological Measurements

A set of neurophysiological sensors that will estimate the ATCOs’ mental states in real time while dealing with the ATC simulation. To achieve this capability, the ATCOs’ brain (Electroencephalogram – EEG) and skin (Electrodermal activity – EDA) conductance will be acquired during the task execution by using wearable, wireless, non-invasive and reliable sensors. In particular, the device for the EEG data collection is the Mindtooth Touch system (https://mindtooth-eeg.com/) with 8 water–-based

electrodes (5 frontal and 3 parietal channels) and Bluetooth low-energy (BLE) connectivity. This system has been used in several real contexts as described in these works [26][27][28][29][30][31][32].

For the EDA data collection, the Shimmer3 GSR+ (https://shimmersensing.com/product/shimmer3-gsr-unit/) or the Research Ring from Biopack (https://www.biopac.com/product/research-ring) will be used. In any case, the selected technology will guarantee reliable signal quality and comfort to be compliant with both ATC settings and validation requirements.

The following neurometrics will then be calculated through offline analysis.

Workload

The workload neurometric, has been calculated as the ratio between the EEG theta activity over the frontal channels, and the alpha EEG activity calculated over the parietal sites.

$$WL = \frac{Theta(Frontal\ Channels)}{Alpha(Parietal\ Channels)}$$

Mental Stress

The mental stress neurometric, has been calculated by using the high beta activity, over the left and rights parietal channels:

$$Stress = BetaHigh\ (P3, P4)$$

Vigilance

Increased frontal activity in the beta band, more in right than in the left hemisphere, is correlated with vigilance decrement [33]. Therefore, the vigilance neurometric was defined as:

$$Vigilance = -Beta(AF4, AF8)$$

Arousal

The arousal (emotional stress) neurometric, has been associated and calculated by using the Tonic component of the EDA (or GSR) signal. Whilst the mental stress neurometric is most related to the cognitive (instantaneous) effect that stressful event may induce, the EDA-based metric, shows the effect that the stress may induce in autonomic system (most related to emotional variations), that has been demonstrated to persist over time [34].

Approach Withdrawal

Finally, the approach-withdrawal neurometric, related to the level of acceptance experienced by the user in front of a specific operational solution, has been calculated by the difference between the EEG alpha activity over the frontal rights sites, and the EEG alpha activity over the frontal left sites.

$$AW = Alpha\ (Frontal\ Dx) - Alpha\ (Frontal\ Sx)$$

The concomitant variation of the previous ascribed neurometrics will be also used to compute a measure of the human performance envelope (HPE) of the operator during the operational task. Such model will consider co-variations both within each HF (e.g. Low vs High Stress) and between different HFs (e.g. Vigilance vs Workload), to consider their simultaneous coexistence [35].

5.1.8.3 Analysis methods

For statistical analysis, all results from the Likert Scale questionnaires will be analysed through post-hoc tests to identify potential correlations and dependencies between the different dimensions of Trust, Workload, Situational Awareness, and Usability. As Likert Scale questionnaires are considered ordinal data, the assumption that the data follows a normal distribution is violated, therefore ordinal logistic regression or Wilcoxon Rank test will likely be used for analysis.

Table 14 summarises the variables assessed, the assessment and data gathering method, when the data gathering will be performed, and the data analysis envisioned.

Table 16: Variables assessment summary

Variable	**Assessment method**	**Timing of the assessment**	**Data Analysis**
Workload	Workload rating scale	After each condition	Ordinal logistic regression
	EEG	During the experimental task performance	z-score on Neurometrics
Stress (cognitive and emotional)	EEG/EDA	During the experimental task performance	z-score on Neurometrics
Vigilance	EEG	During the experimental task performance	z-score on Neurometrics
Acceptance (Approach with-drawal)	EEG	During the experimental task performance	z-score on Neurometrics
Situational Awareness	SA Rating Scale [21]	After each condition	Proposed by the author
Shared Situational Awareness	Shared and complementary situation awareness for human and agent [20].	Post-hoc comparison	Ordinal logistic regression
Task Performance	SME Observational Grid / Rating Form [21]	During the experimental task performance	Ordinal logistic regression

Usability	UX Acceptance Index	After each condition	Ordinal logistic regression
Situational Trust	Situational Trust Scale for Automated Driving (adapted)	After each condition	Ordinal logistic regression
Dispositional Trust	Automation-induced complacency potential – revised	Before the familiarisation with the tool	Ordinal logistic regression
Shared Mental Model Trust	System Causability Scale	After each condition	Ordinal logistic regression
Human Error	Switch log data SME Observational Grid / Rating Form [21]	During the experimental task performance	Ordinal logistic regression
Timeliness	Switch log data SME Observational Grid / Rating Form [21]	During the experimental task performance	Ordinal logistic regression

5.1.9 Exercise planning and management

5.1.9.1 Activities

The following activities Table 17 will occur across the period of the project leading up to and after the validation exercise.

Table 17 Activities involved in achieving the validation exercise

Phase	Activities	Responsibilities
Preparation	Development of the validation plan; preparation of the scenarios in the experimental platform (selection of traffic samples; creation of scenarios); creation of simulation sessions, questionnaires, and debriefing guidelines; selection of controllers; creation of scenarios guidelines for controllers, preparation of training and familiarization sessions; generation of ethics, data protection and informed consent paperwork.	Validation Lead (Deep Blue) and Simulation and Ethics Assessment Lead (ENAC)
	Development of the mental and task models mock-up; calibration of neurophysiological sensors and creation of questionnaires.	Validation Team (Neurophysiological assessment lead UNIROMA1)

Exercise	Execution of the training and familiarization session, execution of the simulation exercises and data gathering	Validation Lead / Validation Team
Post-Exercise	Analysis of data; reporting of findings. Questionnaires and models update.	Validation Lead / Validation Team (UNIROMA1 for neurophysiological measures and correlation analysis)

5.1.9.2 Roles and responsibilities in the exercise

This sub-section describes the roles and responsibilities of the participants involved in preparing, conducting and analysing the exercise.

Validation Lead

The Validation Lead for the TRUSTY project is Deep Blue through WP6. The responsibilities of the Validation Lead are to ensure activities are defined, planned and conducted, in a scientifically rigorous and ethical manner, to meet the project validation objectives. This will be achieved through several mechanisms including:

- Definition of a validation plan for testing and validating outputs developed in WP4 and WP5;
- Ensure the assessment highlights performance benefit and the operational feasibility of the TRUSTY concept through qualitative and quantitative measurements;
- Defines validation objectives and success criteria which, once tested, will generate recommendations and requirements for future development of the TRUSTY concept;
- Elicits requirements for the platform/tools to be used for the validation.

Simulation Lead

An essential and valuable part of the validation exercise is access to an appropriate test platform which can achieve a high level of ecological fidelity of the real ATCO operating environment. Through project consortium members, the project is privileged to have access to the ENAC ACHIL facilities in Toulouse. Thus, ENAC are appointed as the Simulation Lead to establish the outputs from WP4 and WP5 in the simulator facility for use in the validation exercise, and moreover to co-ordinate the validation exercise during its conduct. This will be achieved through several mechanisms including:

- Enabling a small-scale experiment with human participants on a dedicated platform;
- Providing software and hardware required for such experiments;
- Recruitment and management of ATC operators as experimental subjects through an ethically sound mechanism;
- Ensure operational scenarios can be performed in which the effectiveness of a transparent and explainable system for RDTs will be investigated.

Human Performance Measurement

Human performance measurement is the responsibility of Deep Blue through WP6 and La Sapienza through WP5, as Human Behaviour and Performance SMEs. This requires the down-select, integration, administration and analysis of all tests related to human behaviour and performance in the validation

exercise. This includes the identification of indicators to assess the human performance and the user acceptance quantitatively, qualitatively, objectively or subjectively e.g., performance observations, self-report questionnaires, neurophysiological measures, qualitative interviews, etc.

Ethics, Data Management and Informed Consent

Responsibility for ethics, data management and informed consent is further described in the TRUSTY deliverables D2.4 Data Management Plan [36], D2.1 Report on ethical framework for handling personal data and sharing and access to data within TRUSTY [37], D2.2 H - Requirement No. 1 [23] and D2.3 PD - Requirement No. 2 [38]. Albeit these processes will be overseen by the Validation Lead and Project Coordinator (MDU) to ensure effective implementation and conduct.

5.1.9.3 Time planning

Table 18 Exercise time planning.

Activity	Project Month											
	13	14	15	16	17	18	19	20	21	22	23	24
Preparatory												
Execution												
Data Analysis												
Exercise report writing												
ERR Delivery												

Green line represents the delivery of the TRUSTY concept (visualisation and AI) for test

5.1.9.4 Identified risks and mitigation actions

This section lists the risks associated with the validation exercise, their severity and likelihood and the corresponding mitigation actions.

Table 19 Exercise risks and mitigation actions

Risks	Impact (1-low, 2-medium, 3-high)	Likelihood (1-low, 2-medium, 3-high)	Criticality (calculated based on likelihood and impact)	Mitigation actions
Acquisition devices are prone to record artefacts in the different experiments	3	1	2	Ad-hoc algorithms will be used developed to ensure the possibility to properly process the biosignals in both laboratory and realistic conditions.
Suggested experimental design and analysis are	3	3	3	Periodic revision of planned experimental design and analysis will be done. Also, pilot tests will be always performed, and

not able to cope with research objectives				participants will become confident in the setup, before participating in experiment.
The quality of video and audio feeds provided by the project partner may be inadequate or inconsistent	3	1	2	The project partners will rigorously test data feeds before the experiment to verify quality and consistency
Challenges in integrating the AI tool in the remote digital tower simulator may lead to technical glitches	2	2	2	The project partners will perform an integration testing, being also ready to provide technical support during the experiment
Lack of system user-friendliness might increase workload and stress levels out of researchers' control	1	1	1	Researchers will provide initial training and familiarization with the tool
Due to the simulation environment, ATCOs may become overly reliant on the AI tool, potentially reducing situational awareness out of researchers' control	1	1	1	Researchers will include emphasis on maintaining situational awareness throughout the experiment as per real operations
Assumption of AI 100% accuracy	2	3	2	Researchers will specify AI is not 100% accurate during the initial training and familiarization
The duration of the whole experiment may lead to fatigue affecting performance and feedback	1	1	1	Researchers will ensure the presence of breaks during the experiment
Misinterpretation of correlation between variables	2	1	2	Researchers will use appropriate statistical measures following a rigid methodology for the data analysis

6 References

6.1 Applicable documents

This ERP complies with the requirements set out in the following documents:

Master Planning

[1] European ATM Master Plan. 2020 Edition. SESAR Joint Undertaking, 2019. Print: ISBN 978-92-9216-111-8 doi:10.2829/697407 PDF: ISBN 978-92-9216-110-1 doi:10.2829/10044

[2] Strategic Research and Innovation Agenda (SRIA) Digital European Sky. September 2020. SESAR Joint Undertaking, 2020 (http://europa.eu) Print 978-92-9216-156-9 doi:10.2829/49896 MG-03-20-397-EN-C PDF 978-92-9216-155-2 doi:10.2829/117092 MG-03-20-397-EN-N

[3] Multiannual Work Programme (MAWP) SESAR 3 Joint Undertaking 2022-2031 Print ISBN 978-92-9216-174-3 doi:10.2829/156176 MG-03-21-408-EN-C PDF ISBN 978-92-9216-175-0 doi:10.2829/60154 MG-03-21-408-EN-N

System and service development

[4] DES SESAR maturity criteria and sub-criteria. Edition number 01.01. Edition date 15 Feb 24.

Performance management

[5] DES Common Assumptions. Deliverable ID D4.3, Project Acronym: PJ19-W2 CI, Grant 874473, Call: H2020-SESAR-2019-1. Topic: Content Integration. Consortium Coordinator: EUROCONTROL. Edition date: 29 June 2023 Edition: 00.02.01, Template Edition: 02.00.05.

[6] DES Performance Framework Deliverable ID: D4.4 Project Acronym: PJ19-W2-CI Grant: 874473. Call: H2020-SESAR-2019-1. Topic: Content Integration. Consortium Coordinator: EUROCONTROL Edition date: 29 June 2023 Edition: 00.01.04 Template Edition: 02.00.05

Validation

[7] DES HE requirements and validation / demonstration guidelines. SESAR Joint Undertaking. Edition Date: 15 September 2023 Edition: 03.00

Human performance

[8] SESAR Human Performance Assessment Process TRL0-TRL8 Deliverable ID: PJ19 D4.0.070 Dissemination Level: Public Project Acronym: PJ19 CI Grant: 731765 Call:[H2020-SESAR-2019-1] Topic: Content Integration Consortium Coordinator: EUROCONTROL Edition date: January 2024 Edition: 00.04.02 Template Edition: 03.00.01

Project and programme management

[9] Grant Agreement Project number: 101114838. Project name: Trustworthy Intelligent System for Remote Digital Tower. Project acronym: TRUSTY. Call: HORIZON-SESAR-2022-DES-ER-01. Topic: HORIZON-SESAR-2022-DES-ER-01-WA1-7. Type of action: HORIZON JU Research and

Innovation Actions. Granting authority: SESAR3 Joint Undertaking. Project dates: 1 September 2023 - 28 February 2026.

[10] SESAR 3 Joint Undertaking Project Handbook– Programme Execution Framework, Edition date 11 April 2022, Edition: 01.00.

6.2 Reference documents

[11] P. Ortner, R. Steinhöfler, E. Leitgeb, and H. Flühr, 'Augmented Air Traffic Control System—Artificial Intelligence as Digital Assistance System to Predict Air Traffic Conflicts', AI, vol. 3, no. 3, Art. no. 3, Sep. 2022, doi: 10.3390/ai3030036.

[12] M. Cocchioni, S. Bonelli, C. Westin, A. Ferreira, and N. Cavagnetto, 'Guidelines for Artificial Intelligence in Air Traffic Management: a contribution to EASA strategy', in Neuroergonomics and Cognitive Engineering, AHFE Open Acces, 2023. doi: 10.54941/ahfe1003008.

[13] M. Branlat, P. H. Meland, T. E. Evjemo, and A. Smoker, 'Connectivity and resilience of remote operations: insights from air traffic management', REA Symposium on Resilience Engineering Embracing Resilience, Nov. 2019, doi: 10.15626/rea8.15.

[14] L. Axon et al., 'Securing Autonomous Air Traffic Management: Blockchain Networks Driven by Explainable AI'. arXiv, Apr. 27, 2023. doi: 10.48550/arXiv.2304.14095.

[15] M. Jameel, L. Tyburzy, I. Gerdes, A. Pick, R. Hunger, and L. Christoffels, 'Enabling Digital Air Traffic Controller Assistant through Human-Autonomy Teaming Design', in 42nd IEEE/AIAA Digital Avionics Systems Conference, DASC 2023, Barcelona, Spain: IEEE, Oct. 2023. Accessed: Feb. 07, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10311220

[16] Deliverable number D3.2: Report on the gap analysis including KPIs-KVIs and the development-technical work plan. Deliverable ID: D3.2. Project acronym: TRUSTY. Grant: 101114838 Consortium coordinator: MALARDALENS UNIVERSITET Edition date: 29 June 2024. Edition: 00.01.00

[17] L. Jiang, P. Yang, X. Ma, H. Yang, T. Li, and J. Yang, "Comparison of Detection Technology for Runway Incursion Prevention in Airport Hot Spot," J. Phys.: Conf. Ser., vol. 1570, no. 2020, p. 012052, Jun. 2020, doi: 10.1088/1742-6596/1570/1/012052

[18] Guidance Material on remote airport air traffic services, EASA. Issue 2, February 2019. https://www.easa.europa.eu/en/document-library/acceptable-means-of-compliance-and-guidance-materials/gm-remote-tower-operations-0

[19] Artificial Intelligence Roadmap 2.0 Human-centric approach to AI in aviation. EASA. May 2023. Version 2.0 easa.europa.eu/ai

[20] Cain, A. A., Edwards, T., & Schuster, D. (2016). A Quantitative Measure for Shared and Complementary Situation Awareness. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 60(1), 1823-1827. https://doi.org/10.1177/1541931213601416

[21] Endsley, M., Sollenberger, R., Nakata, A., Stein, E., Situation Awareness in Air Traffic Control: Enhanced Displays for Advanced Operations. Federal Aviation Administration Technical Center. Technical note. ADA375375 Year 2000

[22] Nielsen, Jakob, and Landauer, Thomas K.: "A mathematical model of the finding of usability problems," *Proceedings of ACM INTERCHI'93 Conference* (Amsterdam, The Netherlands, 24-29 April 1993), pp. 206-213.

[23] D2.2 H - Requirement No. 1 Procedures to identify/recruit research participants in TRUSTY. Deliverable ID: D2.2 Project acronym: TRUSTY Grant:101114838 Consortium coordinator: MDU Edition date: 25 July 2024 Edition: 00.01.00

[24] AEON (2022). D4.1. AEON ER Description of the first validation platform. SESAR JU. H2020-SESAR-2019-2.

[25] Razin, Yosef & Feigh, Karen. (2023). Converging Measures and an Emergent Model: A Meta-Analysis of Human-Automation Trust Questionnaires.

[26]Evaluation of a new lightweight EEG technology for translational applications of passive brain-computer interfaces N Sciaraffa, G Di Flumeri, D Germano, A Giorgi, A Di Florio, G Borghini, Frontiers in Human Neuroscience 16, 901387

[27]Ronca, V., Brambati, F., Napoletano, L., Marx, C., Trösterer, S., Vozzi, A., ... & Di Flumeri, G. (2024). A Novel EEG-Based Assessment of Distraction in Simulated Driving under Different Road and Traffic Conditions. Brain Sciences, 14(3), 193.

[28]Borghini, G., Giorgi, A., Ronca, V., Mezzadri, L., Capotorto, R., Aricò, P., ... & Babiloni, F. (2023, November). Cooperation and mental states neurophysiological assessment for pilots' training and expertise evaluation. In 2023 IEEE International Workshop on Technologies for Defense and Security (TechDefense) (pp. 77-82). IEEE.

[29]Di Flumeri, G., Giorgi, A., Germano, D., Ronca, V., Vozzi, A., Borghini, G., ... & Aricò, P. (2023). A neuroergonomic approach fostered by wearable EEG for the multimodal assessment of drivers trainees. Sensors, 23(20), 8389.

[30]Ronca, V., Uflaz, E., Turan, O., Bantan, H., MacKinnon, S. N., Lommi, A., ... & Borghini, G. (2023). Neurophysiological Assessment of An Innovative Maritime Safety System in Terms of Ship Operators' Mental Workload, Stress, and Attention in the Full Mission Bridge Simulator. Brain Sciences, 13(9), 1319.

[31]Pugh, J., Goman, M., Abramov, N., Borghini, G., De Vissche, I., Granger, G., ... & Rooseleer, F. (2023). Air Vortex Upset Prevention & Recovery Training (UPRT) Flight Simulation with Wake Vortex Encounter Events.

[32]Borghini, G., Ronca, V., Aricò, P., Di Flumeri, G., Giorgi, A., Bonelli, S., ... & Babiloni, F. (2023). Teamwork objective assessment through neurophysiological data analysis: a preliminary multimodal data validation. In Neuroergonomics and Cognitive Engineering (Vol. 102).

[33]Molina, E., Sanabria, D., Jung, T., and Correa, A. Electroencephalographic and peripheral temperature dynamics during a prolonged psychomotor vigilance task. Accident Analysis & Prevention, Volume 126, 2019, Pages 198-208, https://doi.org/10.1016/j.aap.2017.10.014.

[34] Borghini, G., Di Flumeri, G., Aricò, P. *et al.* A multimodal and signals fusion approach for assessing the impact of stressful events on Air Traffic Controllers. *Sci Rep* **10**, 8600 (2020). https://doi.org/10.1038/s41598-020-65610-z

[35]P Aricò, G Borghini, G Di Flumeri, S Bonelli, A Golfetti, I Graziani, S Pozzi. Human factors and neurophysiological metrics in air traffic control: a critical review IEEE reviews in biomedical engineering 10, 250-263

[36]TRUSTY Initial Data Management Plan Deliverable ID: D2.4 Project acronym: TRUSTY Grant: 101114838 Consortium coordinator: MALARDALENS UNIVERSITET Edition date: 24 November 2023 Edition: 00.01

[37]Report on ethical framework for handling personal data, and sharing and access to data within TRUSTY Deliverable ID:D2.1 Project acronym: TRUSTY Grant: 101114838 Consortium coordinator: MALARDALENS UNIVERSITET Edition date: March 2024

[38]D2.3 PD - Requirement No. 2 Processing of Personal Data. Project acronym: TRUSTY Grant: 101114838 Consortium coordinator: MALARDALENS UNIVERSITET Edition date: July 2024