SESAR JOINT UNDERTAKING

D3.1:Report on Definition, Specifications and SotA of Artificial Intelligence in Remote Digital Towers

Deliverable ID:	D3.1
Project acronym:	TRUSTY
Grant:	101114838
Call:	HORIZON-SESAR-2022-DES-ER-01
Topic:	HORIZON-SESAR-2022-DES-ER-01-WA1-7
Consortium coordinator:	MDU
Edition date:	09 April 2024
Edition:	00.02.00
Status:	Final
Classification:	PU

Abstract

The study defines the possible definition, specifications for the taxiway and runway monitoring and decision support in RDTs application domain and consequently the functional requirements of the XAI, human-centred XAI, HMI and HAIT solutions in a customized way. Here, it identifies the transparency in AI based on a systematic literature review on AI explainability, HMI and GUI with human-centred XAI in the domain of RDTs, i.e., taxiway and runway monitoring. The task will also identify the SotA techniques and approach for interactive data visualisation, human-centric AI model development and hAIi interfaces, and HAIT in RDT domain.

Authoring & approval

Author(s) of the document

Organisation name	Date
MDU	2023-09-18
ENAC	2023-12-18
UNIROMA1	2024-02-08

Reviewed by

Organisation name	Date
Deep Blue	2024-02-26
ENAC	2024-02-23
UNIROMA1	2024-02-22
MDU	2024/02/26

Approved for submission to the SESAR 3 JU by[1]

Organisation name	Date
MDU	2024-02-07
ENAC	2024-02-07
UNIROMA1	2024-02-14
Deep Blue	2024-02-14

Rejected by[2]

Organisation name	Date

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

[2] Representatives of the beneficiaries involved in the project

© –2023– SESAR 3 JU

Document history

Edition	Date	Status	Company Author	Justification
00.00.01	18/09/23	Initial template	MDU	Initial table of contents and structure of the report
00.00.01	17/10/23	Initial Draft	MDU	Add contents to Section 3 and section 4
00.00.02	06/11/23	Draft	MDU	Updated contents of Section 3 and section 4
00.00.03	23/11/23	Draft	MDU	Add contents to Section 2
00.00.04	18/12/23	Draft	ENAC	Add content to section 3
00.00.05	05/01/24	Draft in New template	MDU	Updated structure, move to new template and add text of section 1
00.00.06	10/01/24	Draft	ENAC	Updated Section 2 and section 3
00.00.07	15/01/24	Draft	ENAC	Updated the section 3
00.00.08	26/01/24	Draft	ENAC	Updated the section 2
00.00.09	07/02/24	Draft	MDU	Updated the section 6
00.00.10	08/02/24	Draft	UNIROMA1	Updated the section 5
00.00.11	14/02/24	Draft	MDU	Formatting and synchronize all references
00.00.12	16/02/24	Draft	MDU	Updated with internal review
00.00.13	19/02/24	Draft	MDU	Ready for external review
00.00.14	28/02/24	Draft	MDU & ENAC	Revised based on the internal review feedback
00.00.15	29/02/24	Draft	MDU	Adding Executive Summary
00.01.00	29/02/24	Final	MDU	Ready for submission
00.01.01	04/04/2024	2nd Round Draft	MDU	Addressing SJU's feedback
00.01.02	05/04/2024	2nd Round Draft	MDU	Final revision
00.02.00	09/04/2024	Final	MDU	Ready for submission

© –2023– SESAR 3 JU

Copyright statement © (2024) – (TRUSTY Consortium). All rights reserved. Licensed to SESAR 3 Joint Undertaking under conditions.

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

© –2023– SESAR 3 JU
EUROPEAN PARTNERSHIP
Co-funded by the European Union

1 Executive Summary

The TRUSTY —Trustworthy Intelligent System for Remote Digital Tower project exemplifies this innovation, aiming to equip RDTs with AI to match the safety and efficiency of conventional ATC towers. By focusing on the trustworthiness of AI systems, the project seeks to ensure that these advanced towers can reliably monitor critical areas such as runways and taxiways. The project will integrate sophisticated visualization techniques into the operator interfaces, enhancing human-AI interaction.

Remote digital towers (RDTs) are a ground-breaking development in ATM/ATC, employing advanced technology to enhance efficiency and safety by allowing ATCOs to operate remotely. RDTs use a network of high-resolution cameras and sensors around airports, offering improved safety, cost reduction, increased flexibility, and a comprehensive view of airport operations. RDTs improve safety by providing a 360-degree view of airports, reducing blind spots, and saving costs by eliminating the need for traditional towers, but they also provide detailed insights into aircraft movements and weather conditions, with advanced image processing and AI algorithms enhancing situational awareness and hazard detection.

AI and related technologies in ATM/ATC, especially in RDT contexts, are revolutionizing the field by enhancing efficiency, reducing workload, and improving overall operational safety. These advancements are not only beneficial in current operations but also hold significant potential for the future training and development of ATCOs due to the human-AI integration. These different studies highlight the challenges and innovative solutions in RDTs, covering machine learning, augmented reality, and remote sensing. They mark a shift towards enhanced safety, improved real-time decision-making, and understanding of digital and autonomous systems in high-stakes environments.

In Operational Scenarios in Remote Digital Towers and the Role of XAI, presents a series of operational use-case scenarios within Remote Digital Towers (RDT) to understand better the needs of Air Traffic Control Officers (ATCOs) and Remote Tower Operators (RTOs). These scenarios are instrumental in showcasing the multifaceted applications and indispensable value of Explainable Artificial Intelligence (XAI) in Air Traffic Management (ATM) and Air Traffic Control (ATC) as part of the TRUSTY project. Through detailed narratives, the section illustrates how XAI can significantly enhance decision-making processes, crisis management, and operational resilience in various challenging situations. The operational scenarios discuss the use of XAI during specific ATC challenges which include crisis management, the impact of adverse conditions, and the integration of Unmanned Aerial Vehicles (UAVs). By providing ATCOs and RTOs with transparent, understandable, and actionable insights, XAI aims to introduce more resilience, efficiency, and willing collaboration between the end user and the AI.

Trusted Intelligent System, provides an in-depth analysis of the integration and impact of AI in Air Traffic Management (ATM), focusing on robust and interpretable AI, Human-Centric XAI and Human Machine Teaming.

The **Robust and Interpretable AI** section underscores the critical importance of developing AI systems that are both robust and interpretable, particularly within the context of the RDT. It highlights the necessity for AI to be trained on diverse data sets for enhanced reliability and to provide clear,

© –2023– SESAR 3 JU

understandable explanations of its decisions recommendations, fostering trust in high-stakes fields such as healthcare, finance, transportation, and air traffic management.

This section discusses the challenges and solutions for implementing Multimodal Machine Learning (MML) to integrate data from various sources for optimization of AI design, emphasizing the importance of ethical considerations. It also addresses the need for AI systems to be resilient against errors and adaptable to new data.

Accountability, transparency, and fairness are identified as key pillars for ethical AI development, ensuring that AI decisions can be audited and are free from bias, thus encouraging user trust. The section proposes a Learning Assurance Process to validate AI tools, like conflict detection in air traffic control, ensuring they are safe and effective.

Human-Centred Explainable AI introduces the pivotal role of human-centric XAI within the TRUSTY project, underlining the project's commitment to developing AI systems that are not only advanced but also transparent, understandable, and, most importantly, trustworthy. Human-centric XAI aims to bridge the gap between AI's complex mechanisms and the user's need for clear, actionable insights. TRUSTY aims to meet users' needs by providing explanations that enhance trust and improve decision-making capabilities, whilst not increasing workload and cognitive demands.

A critical focus of TRUSTY is the methodological exploration and evaluation of XAI techniques. This involves an innovative approach to assessing how users perceive AI explanations in terms of their usefulness, satisfaction, understandability, and the overall performance of the AI system. These human-centric evaluation methods are essential for ensuring that the AI explanations meet the intended objectives of enhancing user trust and acceptance and facilitating more effective interaction with AI systems.

Moreover, the section acknowledges the inherent challenges in maintaining explainability within increasingly complex AI systems. The project will balance achieving high accuracy and maintaining the transparency necessary for user trust and understanding. This involves a strategic focus on specific methods that can effectively assess and improve the explainability of AI systems without compromising their performance.

Human Machine Teaming delves into the dynamics of Human-AI Teaming (HAIT), defining it as the cooperative interaction between humans and AI to achieve shared goals. This collaboration brings significant benefits, including enhanced decision-making, improved problem-solving, and increased creativity across various domains such as the military, healthcare, and ATM.

Despite the evident advantages, HAIT faces numerous challenges spanning design, interaction, social, behavioural, ethical, societal, integration, and coordination. These include uncertainties about AI capabilities, managing complex AI outputs, fostering trust and confidence, addressing human emotions, ensuring fairness, mitigating bias, preventing unintended consequences, and optimising teamwork and human-machine interaction.

A focal point of this section is the exploration of state-of-the-art HAIT within RDT. It emphasises the importance of advanced systems equipped with fail-safe mechanisms, AI technologies, consideration of human factors, and the development of user-friendly interfaces. In RDTs, AI technologies such as speech recognition are highlighted as a potential way of enhancing air traffic controller capabilities.

© –2023– SESAR 3 JU

Co-funded by the European Union

Understanding cognitive demands and workload is deemed essential, alongside ensuring that interfaces are designed to maximise effectiveness and user experience.

In summary, Chapter 4 highlights the integration of AI in RDT, focusing on developing robust, understandable, and ethically sound AI systems. This chapter underscores the necessity of balancing AI's technological advancements with user trust and ethical development for effective air traffic management in an AI-enhanced future.

Human Factor and Cognitive Assessment, delves into the challenges of integrating AI into human teams, emphasising the importance of understanding human factors such as mental workload and stress. It introduces the use of neurophysiological measures, such as electroencephalograms (EEG) and electrocardiograms (ECG), as superior methods for gauging these factors compared to traditional subjective assessments or performance metrics. Specifically, Brain-Computer Interfaces (BCIs) are highlighted for their potential to significantly improve trust and effectiveness within HAIT by offering a direct channel for assessing and adapting to the operator's mental and emotional states, including workload, stress, and trust levels.

A pivotal initiative discussed is the TRUSTY project, which employs passive BCIs to monitor and adapt AI behaviour in response to real-time assessments of an operator's mental and emotional states. This approach not only aims to enhance collaboration but also addresses the challenge of unconscious bias towards AI by utilising neurophysiological measures to provide deeper insights into human-AI interactions. The application of neuroesthetics in artistic domains further enriches our understanding of these interactions, offering valuable perspectives on bias and collaboration dynamics.

In conclusion, underscores the critical role of neurophysiological measures in advancing HAIT. By leveraging these innovative techniques, we can improve the trust, efficiency, and overall effectiveness of human-AI collaborations, paving the way for a more integrated and harmonious future between humans and AI technologies.

© –2023– SESAR 3 JU

EUROPEAN PARTNERSHIP

Co-funded by the European Union

2 Introduction

Artificial Intelligence (AI) has become a transformative force across various sectors, including aviation, where it promises to redefine air traffic control (ATC) through remote digital towers (RDTs). The TRUSTY —Trustworthy Intelligent System for Remote Digital Tower project exemplifies this innovation, aiming to equip RDTs with AI to match the safety and efficiency of conventional ATC towers. By focusing on the trustworthiness of AI systems, the project seeks to ensure that these advanced towers can reliably monitor critical areas such as runways and taxiways. The project will integrate sophisticated visualization techniques into the operator interfaces, enhancing human-AI interaction. This initiative is grounded in the principles of trustworthy AI, which include robust and interpretable AI, multimodal machine learning, resilience, accountability, transparency, fairness, explainable AI, and effective human-machine collaboration. These principles are essential for building intelligent systems that earn the confidence of users and stakeholders, ensuring that AI decisions are not only accurate but also understandable and fair. The TRUSTY project stands at the forefront of this effort, setting a standard for deploying AI in high-stakes environments like ATC, where trust is crucial.

In this report, we search into the definition, specification, and state-of-the-art (SoTA) of the TRUSTY-Trustworthy Intelligent System for Remote Digital Tower (RDT). Hence, this report aims to provide a comprehensive overview of the current SoTA in AI systems, with a focus on their application to RDTs and the trustworthiness of AI. We will explore these systems' theoretical underpinnings, practical implementations, and future directions, ensuring that the RDT project not only meets but exceeds the expectations of trustworthiness in intelligent systems.

The purpose of TRUSTY is to adjust the transparency level to improve the trustworthiness of AI-powered decisions in RDTs. The advancement of intelligent systems has revolutionized numerous industries, and the aviation sector is no exception. Remote digital towers represent a significant technological leap, offering a fusion of advanced sensors, machine learning, and human expertise to manage air traffic with enhanced efficiency and safety. However, the deployment of such systems necessitates a rigorous framework of trustworthiness to ensure reliability, safety, and user acceptance.

To achieve the project objectives, we need to study operational use case scenarios to enable a better understanding of user needs, i.e., Air Traffic Control Officers (ATCOs) and Remote Towers Operators (RTOs). In this way, some scenarios will be presented that explore the multifaceted applications of explainable AI (XAI) in ATC within the TRUSTY project.

The first scenario presents Jean, an ATCO, using XAI at an RDT during a critical system failure, showcasing the system's utility in crisis management. The second scenario involves Marie at Muret-Lherm Aerodrome, where XAI aids in directing air traffic under reduced visibility conditions. In the third scenario, Raphaël at London City Airport leverages XAI to overcome challenges posed by a power outage, demonstrating its role in enhancing operational resilience. The fourth scenario shifts focus to a regional ATC centre, illustrating the integration of XAI in managing UAVs and their interactions with commercial flight corridors, emphasizing the enhancement of human-machine collaboration. Finally, in the fifth catastrophic scenario at an RDT, Alex, aided by the advanced XAI system, adeptly manages a crisis involving dense weather conditions, UAV and crewed aircraft operations, system failure, and additional hazards such as a medical emergency and runway obstructions, employing transparent

© –2023– SESAR 3 JU

insights from XAI for informed decision-making, ensuring safe operations, and enhancing crisis management through collaborative and adaptive approaches.

These scenarios collectively illustrate the indispensable role of XAI in modern Air Traffic Management (ATM) and ATC, emphasizing its significance in crisis situations, adverse conditions, or UAV integration, thereby charting the future of civil aviation in an AI-driven era.

The concept of a Trusted Intelligent System has many aspects, encompassing robust and interpretable AI, multimodal machine learning (ML), robust and resilient ML, accountability, auditability, transparency, fairness, Human-Centred XAI (HCXAI) and human-machine teaming. Each of these components plays a pivotal role in the development and operation of intelligent systems that can be trusted by users and stakeholders alike.

Robust and interpretable AI is the cornerstone of trustworthy AI systems, ensuring that AI decisions are not solely reliable across a range of conditions but also comprehensible to human operators. Thus, the transparency of AI is particularly critical in the RDT design and implementation, where decisions must be made swiftly and with a clear understanding of the underlying rationale.

Multimodal ML is an essential aspect of trustworthy AI systems, teaching computers to process and synthesize information from various inputs—visual, auditory, textual, etc. This capability is crucial for the RDT project, which relies on a multitude of sensors and data sources to provide a comprehensive view of the taxiway, runway, and airspace.

The foundation of the RDT project consists of robust and resilient machine learning models that are carefully designed to withstand data anomalies and maintain optimal performance. Robustness is crucial, considering the high stakes involved and the profound consequences of any potential failure. Simultaneously, accountability and auditability are essential because they offer a structure for traceability and compliance with legal and ethical requirements, all of which are necessary to maintain public confidence and guarantee regulatory compliance. Furthermore, the principles of transparency and fairness are integral for RDT, as they are instrumental in eliminating biases and guaranteeing that AI systems operate with unequivocal clarity and impartiality, thereby ensuring equitable and unbiased decision-making that treats every aircraft and operator with the same level of objectivity.

HCXAI prioritizes the delivery of explanations that are understandable and actionable, fostering trust and enabling effective human oversight of AI systems. The HCXAI is particularly relevant for the RDT project, where operators must fully grasp AI-generated advice to make informed decisions. Furthermore, human-machine teaming encapsulates the collaborative, constructive collaboration between human operators and AI systems, leveraging their combined strengths to optimize air traffic management in the RDT project.

© –2023– SESAR 3 JU

3 Background and related work

The field of ATM/ATC has significantly evolved to meet the increasing demand for air travel. Key to this evolution is the use of multi-agent coordination techniques in ATM, which decentralizes system responsibilities for improved air traffic flow management [1]. Optimization models and algorithms are developed to enhance the capacity and efficiency of systems, particularly in congested terminal areas [2]. The integration of automation, employing computers and flight management systems, aids human controllers in ATM [3]. Additionally, efficient conflict detection methods are vital for safety and flexibility amidst growing air traffic [4]. The field also investigates arrival flow aggregation effects and the role of complexity metrics in ATM performance, accounting for variables such as weather conditions to thoroughly understand air traffic productivity and efficiency [5], [6]. The evolution and implementation of RDTs in the aviation industry have been explored in the context of the Single European Sky initiative [7]. Feasibility studies and regulatory approvals have been conducted to test the concept of multiple remote towers and assess their potential benefits in improving air traffic management.

Remote digital towers (RDTs) are a ground-breaking development in ATM/ATC, employing advanced technology to enhance efficiency and safety by allowing ATCOs to operate remotely. RDTs use a network of high-resolution cameras and sensors around airports, offering improved safety, cost reduction, increased flexibility, and a comprehensive view of airport operations. RDTs improve safety by providing a 360-degree view of airports, reducing blind spots, and saving costs by eliminating the need for traditional towers, but they also provide detailed insights into aircraft movements and weather conditions, with advanced image processing and AI algorithms enhancing situational awareness and hazard detection.

The advancement in RDTs includes the concept of Remote Tower Operations (RTOs), with Multiple Remote Tower Operations (MRTOs) enabling a single ATCO to manage multiple airports simultaneously using innovative display systems and advanced pan-tilt-zoom (PTZ) cameras [8]. Immersive technologies such as Head Mounted Displays (HMDs) were previously integrated into ATM/ATC to enhance data management and analysis [9]. Innovations in interactive spatial sound and haptics improve ATCO perception and safety, particularly in poor visibility conditions [10]. The implementation of virtual views from the tower through multiple cameras increases the realism and detail of the airport environment for ATCOs [11]. The application of augmented vision in remote towers was explored by [12] emphasizing the need for high-resolution digital panoramas. The authors [13] studied the implementation of eye tracking in augmented reality for RDT while in [14] authors investigated the discriminability of flight manoeuvres in remote tower settings, indicating a potential increase in decision errors in RTOs compared to conventional out-of-windows view. Finally, the development of digital assistants like DiTA in multiple remote towers optimizes ATCOs' workload, enhancing operational efficiency and underscoring the need for trust in these systems [15].

In the field of ATC, particularly focusing on RDTs, AI systems have been instrumental in enhancing operational efficiency, safety, and situational awareness. Recent ATM research, especially in RDT, has shown significant progress in digital technologies and machine learning. The work by [16] highlighted machine learning's effectiveness in aircraft and drone detection and tracking, which is crucial for real-time operations. The authors in [17] examined the impact of digital tower technologies on ATCOs' visual capabilities and their safety implications. One significant advancement is the development of an augmented reality system for remote tower operations that integrates visual spectrum (VS) and

© –2023– SESAR 3 JU

infrared (IR) fusion along with optical tracking. This technology aims to improve efficiency and reduce workload, especially under restricted visibility conditions [18]. Another innovation is the Automated Speech-Based Service Requests (ABSR) system, which supports ATCOs by highlighting recognized callsigns, inputting commands, and feeding digital ATC systems. This system has been shown to reduce workload and improve usability compared to traditional methods without ABSR support [19]. Further research has explored the potential of using multimodal augmentations to increase performance in Single Remote Tower contexts. This includes improvements in controllers' situation awareness and performance under varying operational conditions [10].

Safety performance is one of the important aspects of RDT scenarios. A field study conducted at a large-scale airport examined the safety performance of apron controllers based on digital tower technology [20]. The study addressed the effectiveness of RDTs in enhancing situation awareness and ensuring safe operations. The findings of this study can provide valuable insights into the safety considerations and performance of RDTs in real-world airport environments. Reliability analysis methods for RDT systems have also been investigated [21]. This research focuses on assessing the reliability of remote tower technology, particularly in small airports with low passenger and cargo throughput. Understanding the reliability of RDT systems is essential for ensuring their effectiveness and suitability for different operational environments. Designing remote and virtual ATC centres also involves unique challenges and requires a human systems integration approach. Effective human-system integration requires a thorough understanding of the system in which we operate, and the potential human performance in that system. This involves considering the needs, capabilities and limitations of the controllers to ensure that the system is designed to enhance human performance and situational awareness, and to minimise cognitive workload [22]. To achieve this, the AI systems on which the ATCO will base its decisions must be transparent, understandable, reliable, and trustworthy.

XAI is crucial for making complex AI systems transparent and understandable, fostering trust, acceptance and effective management in various sectors [23], [24]. XAI agent for human-agent interaction explores the role of emotions in cognitive AI agents [25] and emphasizes responsible AI with fairness and accountability [26]. From a historical perspective, XAI evolved from expert systems to advanced machine learning approaches aimed at developing human-understandable systems [27]. XAI applications, such as detecting depressive symptoms in mental health, show its expanding impact [28], supported by social science insights for effective human-like AI explanations [29], [30].

XAI is becoming increasingly pivotal in the field of ATM/ATC and, consequently, in RDT. Particularly, the advancement of XAI in the field of ATM has been reported in a literature review article [31] within the framework of the project ARTIMATION[3] supported by SESAR JU. Including ARTIMATION, there are some other research projects supported by SESAR JU, which are summerisd in Table with their timeline and prominent contributions. Apart from the research projects related to ATM/ATC, another notable advancement was the integration of AI to detect conflicts in air traffic by analysing aircraft surveillance data, thereby augmenting situational awareness for controllers [32]. Due to the critical safety requirements, the design of machine learning systems in ATMs must focus on transparency and user acceptance, underlining the necessity of an update to the regulatory framework for explainability [33]. XAI also plays a crucial role in enhancing the resilience of ATM operations against disruptions caused

[3] https://www.artimation.eu/

© –2023– SESAR 3 JU

by reliance on network infrastructures and remote sensors [34]. Furthermore, blockchain technology and self-learning networking architectures are being integrated with explainable AI to build trust with human stakeholders and optimize ATC [35]. The introduction of digital ATCOs, capable of autonomously performing time-consuming tasks, emphasizes the importance of a human-autonomy teaming interface, supported by explainable AI [36]. These advancements are crucial in making AI-driven decision-making in ATM/ATC, and RDT more transparent, reliable, and acceptable to human operators.

Table 1: Summary of the selected SESAR JU funded research porjects contributing to the developement of XAI in ATM/ATC.

Sl.	Project Name	Timeline	Contribution
1	AISA[4]	2020-06-01 - 2022-11-30	Strategy for providing the necessary information to a specific ATM operational environment (en-route ATC) in order to make them trust the automated system.
2	MAHALO[5]	2020-06-01 - 2022-11-30	AI-based Conflict Detection & Resolution tool with different levels of conformance and transparency.
3	TAPAS[6]	2020-06-01 - 2022-11-30	XAI methods for two operational cases: Conflict Detection & Resolution applied to ATC (tactical), and Air Traffic Flow Management (ATFM) (pre-tactical).
4	ARTIMATION[7]	2021-01-01 - 2022-12-31	Tools for Conflict Detection & Resolution and Delay Prediction with explanation through visualisations.
5	SAFEOPS[8]	2021-01-01 - 2022-12-31	A decision-support tool powered by AI to help ATCOs make complex decisions in the context of go-arounds.

In ATM/ATC, XAI represents a significant step forward in managing complex operations, aiding controllers with sophisticated pattern recognition, predictive analysis with data analytics and visualisation, and decision support. It enhances capabilities and provides deeper understanding of AI-driven decisions, fostering trust and clarity among controllers. The integration of XAI into RDTs and ATC systems exemplifies a commitment to advancing aviation technology while ensuring transparency, understandability, and trustworthiness for human operators.

In summary, AI and related technologies in ATM/ATC, especially in RDT contexts, are revolutionizing the field by enhancing efficiency, reducing workload, and improving overall operational safety. These advancements are not only beneficial in current operations but also hold significant potential for the

[4] https://aisa-project.eu/
[5] http://mahaloproject.eu/
[6] https://tapas-atm.eu/
[7] https://www.artimation.eu/
[8] https://safeops.eu/

© –2023– SESAR 3 JU

future training and development of ATCOs due to the human-AI integration. These different studies highlight the challenges and innovative solutions in RDTs, covering machine learning, augmented reality, and remote sensing. They mark a shift towards enhanced safety, improved real-time decision-making, and understanding of digital and autonomous systems in high-stakes environments.

The TRUSTY project, aligned with SESAR objectives, aims to integrate XAI into RDTs and ATM/ATC systems, marking a significant improvement in the enhancement of aviation safety and operational efficiency. XAI's incorporation into RDTs offers an intuitive, transparent decision-making approach, enhancing controllers' monitoring and management capabilities with precision and insight through high-resolution cameras and advanced sensors.

© –2023– SESAR 3 JU

4 Operational scenarios in Remote Digital Tower

The landscape of ATM has undergone a profound transformation with the advent of RDT technology. As a nexus of advanced sensors, high-resolution cameras, and artificial intelligence systems, RDTs promise unparalleled opportunities for enhancing safety, flexibility, and operational efficiency in aviation. The integration of XAI systems with RDTs stands as a beacon of innovation, promising to revolutionize the way air traffic is managed. TRUSTY project, aligned with the objectives of SESAR, focuses on developing operational use case scenarios that illustrate the transformative potential of XAI in enhancing safety and efficiency in civil aviation operations. The incorporation of XAI-based systems into this framework marks a significant leap forward. XAI's capabilities in pattern recognition, predictive analysis, data analytics and decision-making support offer unprecedented opportunities to augment human expertise. This synergy between human intelligence and machine learning fosters a collaborative environment where safety and operational efficiency are significantly enhanced.

Through meticulously crafted scenarios, this project aims to demonstrate how XAI can assist air traffic controllers in identifying potential safety hazards, optimizing traffic flow, and making real-time, data-driven decisions. The following section will detail operational scenarios which demonstrate the role of XAI in augmenting human expertise within RDTs, offering a glimpse into a future where technology and human skill work in tandem to revolutionize ATC. The scenarios will explore various aspects of civil aviation operations, from routine ATM to handling complex, unforeseen situations.

4.1 Scenario 1: XAI-Driven Crisis Management in RDTs[9]

Part 1: Introduction to the Scenario with XAI Integration

Jean, an experienced air traffic controller, is currently overseeing operations from a Remote Digital Tower (RDT), now enhanced with XAI technology. In this airfield management scenario, Jean encounters an increase in aircraft flow, elevating the workload significantly. This surge presents Jean with unprecedented constraints, such as managing simultaneous take-offs and landings on runways. As the traffic volume intensifies and the queue of aircraft awaiting clearance grows, the role of the XAI system becomes indispensable. Distinct from traditional AI, the XAI system aids Jean by not only presenting data in an interactive and relevant way but also providing transparent explanations for its recommendations. For instance, it outlines why one aircraft is prioritized over another for landing based on factors like fuel levels, weather conditions, and emergency statuses, offering Jean a comprehensive understanding of the decision-making process behind the traffic management. Amidst the escalating complexity of managing the increased air traffic, a critical emergency unfolds: one of the aircraft in the congested airspace urgently declares a medical emergency. The pilot reports a passenger experiencing severe chest pain, suspected to be a heart attack, necessitating an immediate and prioritized landing. This adds a layer of urgency to Jean's responsibilities, compelling a swift adaptation of the current traffic management strategies to accommodate this unforeseen priority. The situation demands not only an expedited landing clearance but also coordination with ground medical

[9] *Advanced Regional ATC Center, 2025.*

© –2023– SESAR 3 JU
EUROPEAN PARTNERSHIP
Co-funded by the European Union

services to ensure that medical personnel are ready at the runway as soon as the aircraft touches down, highlighting the critical nature of quick, informed decision-making in air traffic control.

Part 2: XAI in Advanced Crisis Response

As the situation escalates, the XAI system shifts into an advanced crisis management mode, a capability that sets it apart from traditional ATC towers. It begins an in-depth analysis of the airspace, assessing critical factors such as the exact locations and speeds of aircraft, potential collision courses, and relevant emergency procedures. This process is markedly different from the operations in a standard control tower, where controllers rely heavily on visual observations and less sophisticated data analysis AI-based tools. In contrast to the often manual and visually dependent decision-making in conventional towers, the XAI provides Jean with a logically structured action plan. This includes innovative communication techniques and rerouting options, all explained in a manner that is straightforward for Jean to understand and implement. The key difference here is the level of detail and rationale provided by the XAI system. This transparency allows Jean to quickly comprehend the reasoning behind each action proposed by the AI, establishing a trust in the technology that is not typically possible with the opaquer processes of a standard control tower. Jean's decisions are thus informed by a rich, real-time data analysis, enhancing his ability to manage the emergency efficiently and effectively, a stark contrast to the limitations faced by controllers in conventional air traffic control environments.

Part 3: Navigating the Crisis with XAI's Support

As Jean addresses the critical airspace situation, the XAI system continuously refines its recommendations based on real-time data and sequencing procedures of flights. Jean uses these insights, along with radar data and airport surveillance systems, to make informed decisions. The system’s explainability ensures that Jean fully grasps the logic behind directing aircraft using visual signals and emergency codes. Jean's adept handling of the crisis, supported by the system’s transparent and actionable insights, results in the safe management of all flights, including the emergency landing. The incident underscores the indispensable role of explainable AI in modern ATC, especially in high-pressure situations.

This scenario illustrates the crucial role of XAI in ATC, particularly in managing crisis situations at RDTs. The ability of XAI to provide clear, logical explanations for its recommendations is vital, offering ATCOs like Jean an enhanced level of understanding and trust in AI’s capabilities.

4.2 Scenario 2: XAI-Enabled Decision-Making Enhancement[10]

Part 1: Introduction to Advanced Operations with XAI

Marie, an experienced ATCO, starts her shift at Muret-Lherm Aerodrome, which is now enhanced with TRUSTY RDT technology and integrated with XAI. Dense fog challenges the day, but initially, the RDT's advanced systems provide clear imagery despite the poor visibility. As fog causes condensation on camera lenses, impairing visibility, the XAI system becomes essential. Unlike traditional AI, XAI provides

[10] *Muret-Lherm Aerodrome with TRUSTY RDT Technology, France. 2025*

© –2023– SESAR 3 JU

Co-funded by the European Union

Marie not only with data but also with clear explanations for the aids to decision-making process, crucial in managing air traffic under reduced visibility.

Part 2: XAI in Navigating Reduced Visibility

The XAI system analyses the impaired visibility situation, considering radar data, historical weather patterns and ground operations resolution strategy. It provides Marie with understandable, data-driven instructions and alternative navigational methods, explaining each recommendation's logic and rationale. This transparency in AI's decision-support processes enables Marie to make informed decisions based on a deeper understanding of the AI's suggestions, crucial for managing the simultaneous arrival of multiple commercial flights and several private planes in emergency.

Part 3: Strategic Implementation of XAI Recommendations

As all aircrafts approach simultaneously, the XAI system continues to offer real-time, logical guidance based on conflict resolution strategy procedures. Marie uses this information, along with pilot reports and radar data, to make precise decisions. XAI's ability to explain its reasoning aids her in understanding the optimal landing paths, factoring in wind direction, visibility, and runway availability. With the XAI system's assistance, Marie successfully manages the challenging situation, highlighting the essential role of explainable AI in modern ATC, especially in adverse weather conditions and visibility-impaired scenarios.

This scenario showcases the critical role of XAI in ATC, particularly under adverse weather conditions. XAI's ability to provide clear, logical explanations for its recommendations is vital, offering ATCOs like Marie an enhanced level of understanding and trust in AI's capabilities.

4.3 Scenario 3: XAI-Enhanced Resilience in Aerodrome Operations[11]

Part 1: Introduction to Advanced Operations with XAI

Raphaël, a skilled ATCO at London City Airport, starts his shift in a control tower enhanced with TRUSTY RDT technology, now integrated with explainable AI systems. This integration is designed to optimize routine ATM and provide robust responses to unforeseen challenges. A sudden power outage disrupts the control room, leading to a loss of visual monitoring capabilities. Despite the redundancy systems naturally present in the tower, the electrical incident affected multiple centres suspiciously, leaving only some functional services like XAI-based system but cutting off video camera transmissions from all the runways.

Part 2: Managing the Crisis with XAI-Driven Decision-Making Under Pressure

As Raphaël manages the escalating crisis at London City Airport, a new urgent scenario unfolds. Three aircraft signal their immediate needs: two are civil transport airplanes requiring landing permissions, and one faces a critical fuel shortage, demanding priority landing. This scenario pushes the boundaries of Raphaël's skills and the TRUSTY RDT system, enhanced with XAI technology, to their limits. The XAI system, recognizing the critical nature of the situation, initiates a prioritization protocol. It delivers a

[11] *London City Airport with TRUSTY RDT Technology, England. 2025.*

© –2023– SESAR 3 JU

Co-funded by the European Union

swift evaluation of each aircraft's condition, including fuel levels and the time window before reaching a critical state, utilizing data such as the flight plans provided by airlines and coordination with airport transportation services rather than relying solely on RDT-specific technologies like camera feeds. This approach allows the XAI to propose an optimal landing order that ensures safety for all involved. It clearly explains why the aircraft with the fuel emergency must land first, detailing the strategic landing sequence for the remaining flights.

In the event of a power failure, the XAI system immediately engages backup protocols, tapping into alternative energy sources to maintain critical operations. Even as the usual visual monitoring tools become unavailable, the AI leverages a comprehensive database of flight paths and air traffic control communications to maintain a detailed awareness of the airspace. This information, along with the XAI's ongoing analysis, is relayed to Raphaël through external devices, ensuring he has access to crucial data about aircraft positions and movements without needing direct visual confirmation from cameras. This feature of the XAI system, providing data analytics and visualizations, underscores its utility in scenarios where traditional visual aids might be compromised. The system's ability to explain its recommendations, grounding them in robust data analysis, is vital. This transparency and reliance on a broad dataset, including flight plans and air traffic communications, empower Raphaël to trust and follow the AI's guidance, ensuring decisions are informed, precise, and tailored to the unique challenges of managing air traffic under adverse conditions.

Part 3: Human-AI Synergy in Crisis Management

Raphaël, with the aid of the XAI's insights, coordinates a response. He directs the fuel emergency flight to land immediately, using a runway that the XAI system identifies as most suitable based on current conditions. Simultaneously, he instructs the civil transport airplanes to enter a holding pattern, explaining the situation and the estimated wait time, as calculated by the XAI system. As Raphaël tackles the challenge of managing air traffic under constrained conditions, the XAI system continuously updates its analysis and recommendations. This adaptive approach allows him to make strategic decisions regarding incoming flights and ground movements, ensuring safe and efficient traffic management. Thanks to the combined efforts of Raphaël and the XAI system, all three aircrafts land safely. The fuel emergency flight touches down without incident, and the civil transport airplanes follow in a well-coordinated manner.

This demonstration of human-AI synergy showcases the ability of XAI systems to enhance human decision-making under extreme stress. The situation underscores the importance of XAI in fostering effective human-machine collaboration. Raphaël's ability to comprehend the AI's logic and reasoning enhances his decision-making capabilities, ensuring a high level of operational safety during the crisis. In the debrief, the effectiveness of the XAI system in providing actionable, understandable, and logical guidance under high-pressure conditions is highly commended. Raphaël's ability to seamlessly integrate AI insights into his decision-making process is recognized as a key human factor in the successful management of the situation.

In this scenario, the integration of XAI algorithms into the TRUSTY RDT system exemplifies the future of ATM. XAI's capacity to provide transparent and understandable decision-making processes is vital, especially in complex or crisis situations, ensuring that ATC remains efficient, safe, and adaptable to rapidly evolving challenges.

© –2023– SESAR 3 JU
EUROPEAN PARTNERSHIP
Co-funded by the European Union

4.4 Scenario 4: XAI-Enhanced UAV Integration in Air Traffic Control[12]

Part 1: Enhanced Recognition of Complex Situations with XAI and Emergence of Data Mismatch Challenge

At the advanced regional ATC centre, the integration of XAI within the RDT systems marks the start of a routine yet technologically advanced day. The XAI system, with its capability to provide clear, understandable insights into AI decision-making processes, becomes crucial in managing a mix of flights, but also Unmanned Aerial Vehicles (UAVs). When a group of UAVs dangerously enters a commercial flight corridor, the XAI system not only identifies this deviation but also provides the rationale behind flagging these UAVs as potential risks, along with suggested actions.

The event takes an unexpected turn when the XAI system's data on the UAVs' positions conflicts with that shown by the surveillance cameras. This discrepancy presents a critical challenge, as relying on inaccurate data could lead to unsafe decisions. The XAI system, recognizing this anomaly, initiates a deeper analysis. It starts by cross-referencing the UAVs' past flight data with similar historical patterns and related atmospheric phenomena, aiming to reconcile the data mismatch.

Part 2: XAI-Driven Resolution, Strategic Implementation and Continuous Evolution

Through its comprehensive analysis, the XAI system identifies the root cause of the discrepancy. High precision cameras, despite their advanced capabilities, can sometimes face challenges in accurately recognizing UAVs. UAVs are generally small and can operate at significant altitudes or distances from the camera. Due to their size and the distance, even high precision cameras may struggle to capture enough detail to accurately recognize them, especially if the UAVs are beyond the effective range of the camera's resolution and zoom capabilities. Another point is if the UAVs have coloration or patterns that blend with the background (like the sky, trees, or urban landscapes), it can be difficult for cameras to distinguish them from their surroundings. Low contrast between the UAV and its environment can significantly reduce the camera's ability to detect and recognize them. Additionally, UAVs can move at high speeds and change directions quickly. This rapid movement can cause motion blur in the camera's imagery, making it difficult to maintain a lock on the UAV or to process its image accurately, especially if the camera's frame rate isn't high enough to capture such quick movements clearly. Finally, some UAVs are equipped with technologies designed to evade detection, such as anti-surveillance materials that absorb or deflect radar and certain light frequencies, or electronic countermeasures that can interfere with camera sensors. These technologies can make it challenging for even high precision cameras to detect and recognize the UAVs.

Unfortunately, these different elements played into the situation, making the data transmitted by the video systems unreliable. Consequently, the XAI-system launched a deeper analysis, crossing all the video data available before the event as well as the meteorological data based on machine learning algorithms with explanation of the decision processes. With this newfound understanding, the XAI system recalibrates the UAVs' data, aligning it with the visual feed. The system then updates its

[12] *Berlevag Aerodrome with TRUSTY RDT Technology, Norway. 2026*

© –2023– SESAR 3 JU

Co-funded by the European Union

recommendations for rerouting commercial flights, providing the ATCO with a reliable and accurate understanding of the situation.

Part 3: Post-Incident Review and Future Implications

Equipped with accurate, XAI-verified information, the ATCO effectively manages the airspace, safely rerouting commercial flights and communicating with UAV operators to rectify their course. This incident not only demonstrates the critical role of XAI in complex decision-making but also highlights its importance in continuously evolving and refining ATM protocols and AI algorithms.

Post-incident, the ATCO reviews the performance of the XAI system, appreciating its ability to analyse and resolve unexpected challenges dynamically. This scenario underscores the indispensable role of XAI in future ATM, particularly in scenarios involving UAVs in the frame of visual recognition in RDTs. The continuous development of XAI technology is essential to ensure that ATC remains safe, efficient, and transparent in an increasingly complex airspace use.

This scenario emphasizes the importance of XAI in not only handling routine operations but also in resolving unexpected and complex issues that arise in modern ATM, particularly with the integration of UAVs. The ability of XAI to perform deep, contextual analyses and provide comprehensible, actionable insights is pivotal in maintaining safety and efficiency in the face of technological advancements.

4.5 Scenario 5: Crisis Management in a Remote Digital Tower with UAVs, Weather, and System Failure

Part 1: Adverse Weather Operations and Initial Challenges

During a challenging weather day, with a dense fog enveloping the area, both scheduled UAV flights and crewed aircraft operations persist. The severe conditions drastically reduce visibility, posing significant difficulties for Alex, ATCO responsible for monitoring these flights in the aerodrome. The digital tower's cameras, designed to provide clear images of the airspace, struggle under these conditions, leading to a notable degradation of the approaching landings.

Part 2: Transparent Explanation in Specific Case of System Failure and Escalating Crises

As the weather continues to worsen, a critical crisis emerges within the remote digital tower system. A key component malfunctions, leading to significant disruptions in both communication and data processing. Amidst this chaos, Alex, who oversees the operation and safety of the airspace, is alerted to the system failure. Further complicating matters, several cameras crucial for detecting UAVs become non-operational, creating a substantial surveillance gap.

In this escalating situation, two critical and urgent events unfold. First, an aircraft reports a medical emergency, necessitating an immediate and unplanned landing. This development adds a layer of complexity and urgency to the already challenging scenario, as Alex must coordinate a safe landing path in conditions of reduced visibility and hampered communication.

Simultaneously, a new hazard emerges on the airfield. One of the working cameras detected that a heavy piece of metal, likely dislodged by strong winds, lands on one of the runways. This presents a

© –2023– SESAR 3 JU
EUROPEAN PARTNERSHIP
Co-funded by the European Union

highly dangerous situation, especially since the malfunctioning cameras impair Alex's ability to detect and assess this new hazard effectively. The presence of this metal on the runway, coupled with the reduced surveillance capabilities, heightens the risk of a serious incident, particularly with the insights into how these adverse conditions impact decision-making processes.

These concurrent crises put the digital tower system and Alex to the test. During this critical time, TRUSTY, the advanced XAI system, proves invaluable. It not only informs Alex about the system failure and its implications but also offers essential insights and clear explanations for managing these simultaneous emergencies. This support is crucial for Alex to navigate the compounded challenges of adverse weather, technical failures, and unforeseen dangers, ensuring the highest level of safety and efficiency in air operations.

Part 3: Crisis Management and Post-Crisis Analysis

Faced with communication challenges due to the system failure, Alex struggles to coordinate with both crewed aircraft and UAV operators. In response, TRUSTY suggests alternative communication methods and assists in rerouting aircraft to ensure safety. Alex then initiates emergency protocols, focusing on radio communications and implementing contingency plans for rerouting aircraft. TRUSTY supports these efforts by providing real-time suggestions. Additionally, Alex makes human-in-the-loop adaptations to the flight plans, adjusting altitudes and rerouting, with TRUSTY's assistance, to maintain safe separation between the aircraft.

Throughout the crisis, Alex collaborates with neighbouring ATC sectors and UAV operators, sharing information and managing the airspace collectively. TRUSTY plays a pivotal role in facilitating this exchange of data and insights. It continuously assesses the impact of fog on surveillance, suggesting reliance on secondary systems and additional sensor inputs. After the crisis subsides, Alex and TRUSTY conduct a thorough analysis of the emergency response. They review the decisions made, identify areas for improvement in handling adverse weather conditions, and implement lessons learned to enhance future responses.

Despite the catastrophic scenario involving dense fog and a system failure, Alex, with the assistance of TRUSTY, successfully manages the crisis. The transparent insights from XAI empower the controller to make informed decisions, ensuring the safe operation of both crewed and uncrewed aircraft in challenging conditions. The collaborative crisis management approach enhances communication and coordination, contributing to a resilient and adaptive remote digital tower environment.

4.6 Use-Case conclusions

In conclusion, the TRUSTY project exemplifies the revolutionary role of XAI in enhancing ATC, achieved through a series of carefully designed scenarios. These scenarios, pivotal to TRUSTY, demonstrate how XAI and digital assistants can substantially improve ATMs by helping controllers' pinpoint safety hazards, streamlining traffic flow, and enabling real-time, informed decision-making. Spanning a wide array of civil aviation operations, from everyday management to handling intricate, unexpected situations, they underscore the adaptability of XAI.

A key focus is on bolstering situational awareness, especially in monitoring runways and taxiways, and on the pivotal role of XAI in diminishing the cognitive workload of air traffic controllers. This reduction in workload not only boosts efficiency but also curtails fatigue, enhancing overall operational safety. A

© –2023– SESAR 3 JU

Co-funded by the European Union

central aspect of TRUSTY's methodology is its fluid integration of XAI's digital prowess with the dynamic nuances of ATM, ensuring that human controllers remain integral to decision processes.

This approach accentuates TRUSTY's dedication to marrying technological advances with human-focused considerations. It marks a notable advancement in ATM evolution, aiming to heighten the safety of group aviation operations. The project endeavours to employ more explainable AI systems, fostering a nuanced collaboration between humans and machines. This fine-grained synergy is intended to make complex AI decisions more transparent and understandable to human operators, thereby enhancing trust and effectiveness in high-stakes aviation environments. This integration of XAI is not only crucial for current operations but will also be a cornerstone in the future of civil aviation, ensuring safety, efficiency, and adaptability in increasingly complex airspace environments.

© –2023– SESAR 3 JU

EUROPEAN PARTNERSHIP

Co-funded by the European Union

5 Trusted Intelligent System (TIS)

5.1 Robust and Interpretable AI

Robust and interpretable AI refers to artificial intelligence systems that are designed to be both reliable and understandable. These systems are trained on diverse datasets and are tested to ensure their performance in various scenarios, making them more robust compared to traditional AI models. Additionally, these AI systems provide clear explanations for their decisions and allow users to understand the reason behind the outcomes, which is very important in high-stakes applications like healthcare, finance, and transportation. Interpretable AI can help build trust between humans and machines, enabling more effective collaboration in decision-making.

5.1.1.1 Multimodal Machine Learning

In the academic landscape, Baltrušaitis et al. [37] dive into the topic of Multimodal Machine Learning (MML), which is a branch of Artificial Intelligence. This field is concerned with teaching computers to grasp and connect the information they receive from various senses like seeing, hearing, sensing and text. It's a crucial part of helping computers make sense of the world around us.

On the other side, Parcalabescu et al. in [38] offer a fresh perspective on multimodality in the context of MML. They argue that how we understand multimodality should depend on the specific problem that a computer is trying to solve. Different tasks require different types of information, and the best way to understand this information is in the context of the task at hand. Therefore, their new definition of multimodality is centred around the information that is most relevant for a particular computer task.

Together, these two papers analysed the broad field of MML. Baltrušaitis et al. [37] introduce the field, while Parcalabescu et al. [38] offers a new way to think about it, emphasizing the importance of tailoring the understanding of multimodality to the task at hand.

5.1.1.2 Challenge

In the realm of academic research, the study of Liang et al. [39] focuses on the fascinating challenges posed by multimodal machine learning. This branch of machine learning deals with data coming from different sources like images and text, and it brings about unique and tricky issues for the machine learning community. These challenges are because the data sources are quite different, and there are often connections between them. To tackle these challenges, the paper introduces a system of six main technical problems: representation, alignment, reasoning, generation, transference, and quantification. These challenges cover both the historical aspects of the field and more recent trends. They involve things like figuring out how to model and learn from this diverse data, understanding the different aspects of this diversity, and dealing with the learning and optimization issues that arise when the data is so varied.

A different study [37], explores more practical considerations for the application of MML. This study looks at what can be done with MML, like recognising speech, describing events, answering questions about images, and finding media in large collections. However, the study highlights the technical hurdles to overcome. These challenges are about how to represent and translate the different types of data, align them properly, merge them together, and learn from them, especially when one type of

© –2023– SESAR 3 JU

data doesn't have much information. These challenges are especially important when one of the data sources doesn't have much useful information, like not enough labelled data to learn from.

Together, these two papers shed light on the complexities of MML, showing the academic and practical sides of the field. Liang et al. [39] identify the theoretical challenges, while Baltrušaitis et al. [37] illustrate the real-world applications and the hurdles that must be overcome.

5.1.1.3 Suggestions regarding MML

In [39], the authors introduce a new way of looking at the challenges of multimodal learning, which involves understanding and connecting data from different sources like images and text. They say that these challenges are not studied enough in traditional single-source machine learning and need more attention to advance this field. They also suggest that this new system of challenges will help organize future research and identify the problems that still need solving in MML. Moreover, they call for future research to develop solid theories and practical methods to define and measure the differences and relationships between different data types in multimodal datasets.

The authors of the paper [38] also emphasize the need for better ways to combine different types of data in machine learning. They point out that the current methods are often random and lack a solid foundation. They propose that more research should be directed toward creating well-thought-out methods for combining data types. Additionally, they highlight the importance of ethical considerations in developing multimodal machine learning systems, suggesting that it should be done in a way that aligns with ethical standards in human society.

The paper [39] takes a high level view of the field of MML advancement. They highlight key questions that need addressing in the long term, such as finding better ways to represent and connect data from different sources, transferring knowledge between them, and understanding the complexities of different data types. These questions provide a roadmap for future research in the field of multimodal machine learning.

5.1.1.4 MML in the context of RDTs

The remote digital tower requires planning tools that support the controller with tasks such as sequencing flight movements, rearranging them according to situational demand, and organising remote controller positions, etc. Machine learning models developed to support these tasks can use data coming from different modalities, such as video streaming of visual surveillance, radar signals, communication through text messages, weather reports, and tabular data from air traffic management systems. In MML, these different modalities can be fused and encompassed to provide holistic decisions that can reduce uncertainty and false alarms and improve the trustworthiness of AI systems. Figure 1 illustrates the schematic of MML in the context of RTDs.

© –2023– SESAR 3 JU
EUROPEAN PARTNERSHIP
Co-funded by the European Union

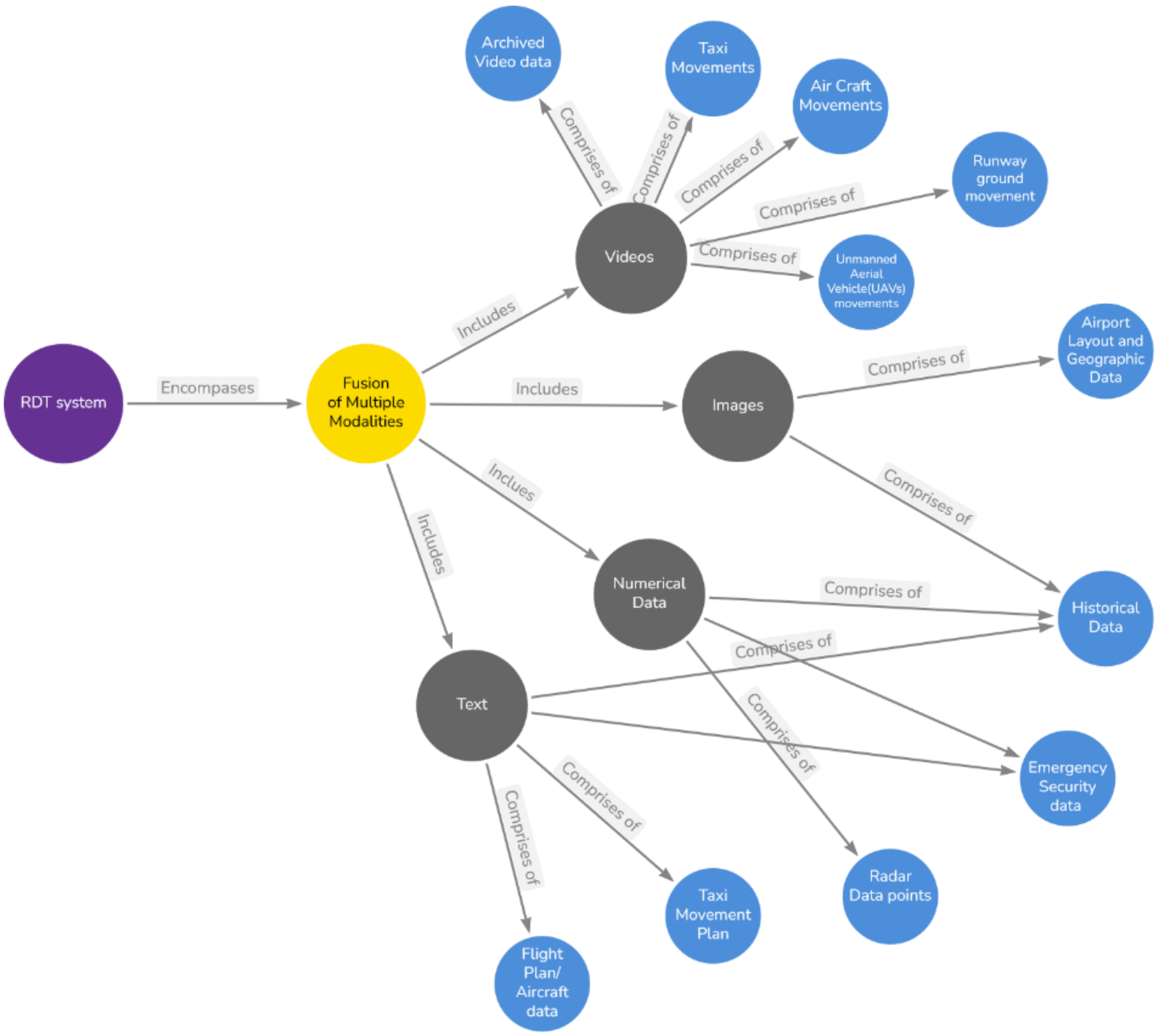

Figure 1: MML in the context of RTDs.

5.1.2 Robust and resilient ML

A good capability of a machine learning model is known as robustness and resilience, where the ML model can resist error, outliers, and data distortion. It also involves the capacity to detect potential issues, make necessary adjustments, and continue functioning effectively when encountering unseen or unexpected noisy data. In crucial application areas like healthcare, banking, and transportation systems, such as ATMs, this robust and resilient ML model plays a vital role in making judgments and predictions. The diagram presented in Figure 2 illustrates the comprehensive process through which machine learning models demonstrate robustness and resilience.

© –2023– SESAR 3 JU
EUROPEAN PARTNERSHIP
Co-funded by the European Union

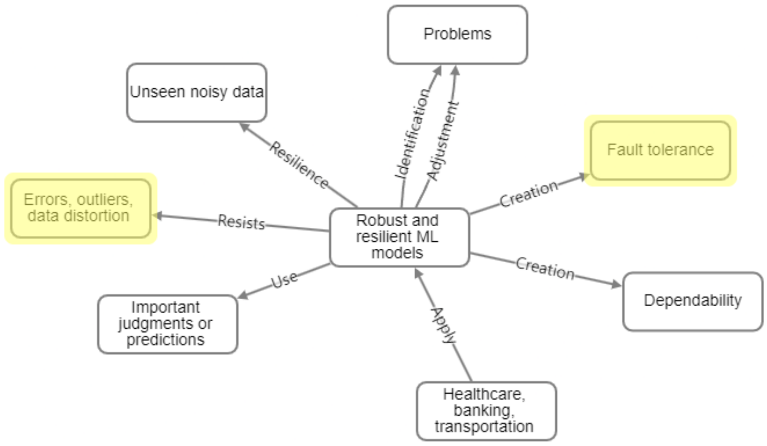

Figure 2: Workflow robust and resilient ML model.

Numerous methods can enhance the robustness of deep neural networks when applied to image classification tasks. In our context, these tasks are essential for detecting objects in remote digital tower systems on runways and taxiways. Some of these techniques include data augmentation, increasing the amount of labelled data, and employing various network architectures such as AlexNet, SqueezeNet, VGG-19, DenseNet-121, and ResNet-50. However, it is important to note that while these techniques have the potential to improve the robustness of image classification, their effectiveness may vary across different cases [40].

5.1.3 Accountability & Auditability

Accountability and auditability (Figure 3) are increasingly important considerations in the development and deployment of artificial intelligence (AI) systems. Ensuring accountability in AI systems means that there is a clear understanding of who is responsible for the actions taken by the system and that there are mechanisms in place to monitor and enforce compliance with ethical and legal standards. Auditability, on the other hand, refers to the ability to assess and verify the performance and behaviour of an AI system, including its decision-making processes and outputs.

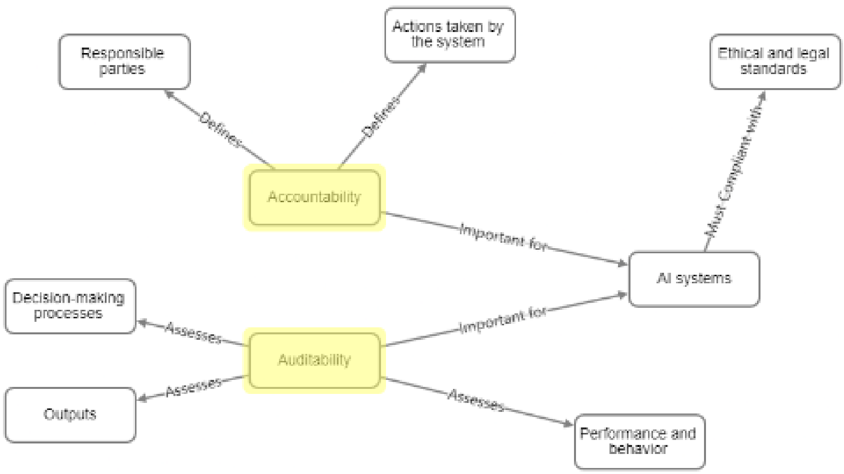

Figure 3: A high-level view of the Accountability and Auditability of ML.

© –2023– SESAR 3 JU

5.1.4 Transparency and Fairness

Transparency and fairness in AI are the core principles and practices that guarantee AI systems operate with clarity, accountability, and impartiality. This encompasses using diverse and representative training data, employing transparent and explainable algorithms, and ensuring a fair and unbiased decision-making process. **Transparency** fosters user trust by allowing them to understand the system's decisions, while **fairness** ensures the system doesn't discriminate based on personal attributes. These principles are essential for building dependable AI systems applicable across various domains like healthcare, finance, criminal justice, and education.

5.1.4.1 Challenges in fairness research

The paper [41] identifies four key challenges in ML model fairness research:

(1) **Balancing the trade-off between fairness and model performance**, where if we don't consider the social and cultural factors when using machine learning, we might end up with models that are unfair, unethical, and even illegal. The machine learning community has come up with different ways to make models fairer, but it's a bit tricky because making a model fair can sometimes make it less good at its job. Sometimes, though, making a model less accurate on purpose is a way to fix unfairness.

(2) **(Dis)agreement and Incompatibility of “Fairness”**, there's a debate in the literature about whether it's more important to be fair to *individuals* or *groups* when using machine learning models. Fairness metrics usually focus on one or the other but not both. Some methods that aim to make things fair for groups can make things worse for individuals within those groups. Also, the way fairness is defined mathematically doesn't always match how society, economics, or the law see it. This makes improving fairness in machine learning challenging. To address these issues, the community needs to find ways to combine different fairness measures and categorize their differences, trade-offs, and preferences, which is a tough task.

(3) **Tensions with Context and Policy**, the current research on fairness in machine learning often tries to make things fair without really understanding the root causes of unfairness. This approach doesn't consider the social and cultural factors that can lead to bias. Instead of just trying to reduce unfairness, we need to pay more attention to the real-life context in which these decisions are made. Also, the data used to train machine learning models often reflects past biases and may not accurately represent the real world. *Researchers should work closely with industry partners to study these models in real-world situations and involve policymakers more in discussions about fairness and standards.* This is a challenging problem, but it's crucial to make progress in the field of fairness in machine learning.

(4) **Democratisation of ML vs the Fairness Skills Gap**, now a days machine learning is more accessible with many tools and cloud-based solutions. This democratization of machine learning can lead to both positive and unintended, socially insensitive uses. However, the challenge is that addressing bias and fairness in machine learning requires expertise, and there's a gap in tools and resources for those without extensive technical knowledge. To keep up with the growing use of machine learning, there's a need for open-source tools, improved educational resources, and comprehensive frameworks to address fairness and bias issues, especially in situations where one model depends on another.

The transparency and fairness of AI-based systems are being ensured through the learning assurance process. This process involves technical development to cover the specific learning processes of AI

© –2023– SESAR 3 JU
EUROPEAN PARTNERSHIP
Co-funded by the European Union

systems. The learning assurance process should provide relevant and understandable information on how the AI application arrives at its results/decisions and perform a safety risk analysis and mitigation process to avoid being considered a black box [42].

Authors of [42] focus on AI systems' learning process in ATC. They tested a conflict detection tool that uses machine learning to predict aircraft separation violations. **The tool uses two methods: estimating minimum distance and identifying potential conflicts.** They used extreme gradient boosting for both methods and then validated the learning process using the EASA's W-shaped methodology. The goal was to understand the tool's performance and avoid the "black-box" effect. **The research found that prediction accuracy decreased as aircraft got closer and that the EASA's methodology needed more time-dependent analysis. AI systems that rely on prediction time or accuracy need more analysis as these factors change over time.** Further study is required to find the relationship between the number of samples and metric deterioration. Lastly, more regulation and certification for AI systems in aviation are needed, and new guidelines for data analysis are necessary to ensure valid performance.

5.1.4.2 A Case Study of Chatbot in Aviation education and research

Recently, [43] presented a case study investigating the impact of ChatGPT (a Chatbot developed by OpenAI) on aviation education and research based on surveys of graduate students of Beihang University, the leading aviation university of China. The survey results reveal that students find it helpful for saving time and efficiency. The major findings of the case study were that students found positive experiences with ChatGPT as they felt they had more access to knowledge and an effective learning environment. Most of the students cross-validate the results with other sources, but interestingly, the female students tend to be even more diligent and critical in this verification process.

The three main tasks that students take help from ChatGPT are programming, state-of-the-art identification, and terminology explanation. Using ChatGPT in aviation faces significant challenges in handling confidential information, conducting innovative research, and processing large volumes of text. Along with those, here are a few important ones:

(1) **Reliability and Over-Reliance:** Using AI like ChatGPT can be risky because it might provide wrong or misleading information. This is a big problem in aviation, where safety is crucial. Students might not question the AI's answers even when they're incorrect.
(2) **Training and Maintenance:** ChatGPT needs constant updates to work well. It takes a lot of time and effort to keep it up to date, which makes it impractical for cutting-edge research.
(3) **Ethical and Bias Concerns:** These systems can learn biases from training data. This could lead to unfair or unethical outcomes. Also, they are vulnerable to cybersecurity threats, and if not adequately protected, they could be hacked, causing problems in aviation.

In short, using ChatGPT in aviation has many challenges that need careful consideration and management by all involved parties.

The students are against completely getting rid of this technology in university education. They think it would be tough to ban it only in some places, like certain countries. They believe that students would try to find ways to still use it because they see it as very useful for their learning. However, there are also challenges and limitations that need to be addressed for a successful and safe application of chatbots in the aviation domain. The paper also suggests future research directions in this area, specially making the chatbot more trustworthy.

© –2023– SESAR 3 JU

5.2 Human-Centred XAI

HCXAI prioritises meeting the needs of users who rely on explanations. To achieve this, XAI explanations should possess key attributes [44]:

(1) They must be understandable, catering to both experts and non-experts in AI. XAI aims to enhance transparency, necessitating explanations with accessible, commonly used terminology that is interpretable and intuitive.
(2) Actionable explanations empower users to make informed decisions when interacting with AI systems. These explanations should enable users to comprehend, trust, and effectively manage the AI system while generating high-quality, interpretable rationales for AI decisions.
(3) Trustworthiness is paramount, demanding accuracy and reliability in explanations. Grounding explanations in real-world examples and ensuring transparency in decision-making are essential to foster confidence among users. Given the inherent opacity of AI models, XAI seeks to alleviate trust challenges by providing explanations that bolster stakeholders' confidence in the utilisation of AI models.

5.2.1 Approaches of Human-Centred XAI

HCXAI involves different approaches that focus on explaining AI system outputs in a way that is understandable and meaningful to humans [45].

One approach to HCXAI is termed as Reflective HCXAI, which takes a socio-technically informed perspective on XAI. It critically reflects on the dominant assumptions and practices of the field and considers the values of diverse stakeholders, especially marginalised groups. Reflective HCXAI aims to propose alternative technologies that are sensitive to socio-organizational contexts and address the limitations of current AI systems [46].

The Human-Centred Artificial Intelligence (HCAI) framework is another approach to Human-Centred XAI. It emphasizes designing AI systems that offer high levels of human control and high levels of computer automation to increase human performance. The framework helps to understand when full human control or full computer control is necessary and how to avoid the dangers of excessive control. The methods of HCAI aim to produce designs that are reliable, safe, and trustworthy [47].

5.2.2 Human-centred evaluation methods

Several literature reviews explored the state of XAI research, highlighting the necessity and challenge of understanding user needs and conducting human-centred evaluations of explainable models [1], [48].

For example, in this study [48], the review discusses the emergence of various terminologies and categorizations for explainable AI (XAI) evaluation methods due to their multidisciplinary nature. It identifies four key categories for HCXAI evaluation: Trust, Explanation Usefulness and Satisfaction, Understandability, and Performance. The study findings suggest that user trust in a model is influenced by both actual and perceived accuracy, emphasise the importance of understandable explanations, highlight a preference for human-generated explanations, and note that an optimal level of transparency is necessary to balance the cost and benefit of explanations.

© –2023– SESAR 3 JU

In another related work by [1], they used similar categories for human-based XAI evaluations (i.e., trust, understanding, usability, and human-AI collaboration performance). The review identified the proliferation of XAI research but emphasized the need for more transparent and comparable human-based evaluations across different application domains. It also underscores the importance of incorporating insights from cognitive and social sciences in XAI research to enhance human-centred approaches and proposes practical guidelines for future studies in this field.

5.2.3 Challenges and Advancements in Human-Centred XAI

One of the main challenges in HCXAI is to evaluate the effectiveness of XAI techniques, which are becoming increasingly complex. The authors in [49], propose to address this challenge by focusing on the explainability of simple mathematical models and assessing how people perceive the comprehensibility of different model representations. This approach allows diverse stakeholders to judge the intelligibility of fundamental concepts that more complex AI explanations are built from.

5.2.4 Human-Centred XAI for Remote Digital Towers (RDT)

The related fields, such as smart cities and industrial asset management, can provide valuable insights and guidelines for the development of XAI in remote digital towers.

Explainable AI for smart cities is a relevant area of research that can provide insights into the development of HCXAI in digital remote towers. The lessons learned from state-of-the-art research in XAI for smart cities can inform the implementation of XAI-driven systems and architectures in the context of digital remote towers. This includes addressing technical challenges, ensuring transparency and accountability, and leveraging XAI techniques to improve decision-making processes [50].

Furthermore, the application of XAI in the prognostic and health management (PHM) of industrial assets can provide valuable insights for human-centred XAI in digital remote towers. The balance between accuracy and explainability is a key consideration in PHM, and similar trade-offs may need to be made in the context of digital remote towers. The involvement of human operators, explanation assessment, and uncertainty quantification are important factors to consider in developing XAI solutions for PHM, and these insights can be applied to the development of HCXAI in digital remote towers [51].

In conclusion, considering the needs and experiences of human operators, prioritizing privacy and trustworthiness, and leveraging explainability techniques, human-centred XAI can enhance the effectiveness and usability of AI systems in digital remote towers.

5.2.5 The Challenges and Opportunities of AI and XAI for ATM

The use of AI in ATM has seen a surge in interest and funding, particularly in the 2010s, driven by the availability of massive volumes of data and the efficiency of computer graphics card processors in accelerating learning algorithms. A recent survey by [31] provides a comprehensive overview of the current state of the art in AI and XAI for ATM. The authors identify several key trends in AI research and development in ATM, including the use of machine learning and deep learning such as Neural Networks (NN), Random Forest (RF), Support Vector Machine (SVM), and others to develop more accurate and efficient algorithms for ATC tasks, as well as the development of XAI methods to make AI systems more transparent and understandable to end users.

© –2023– SESAR 3 JU

Co-funded by the European Union

The authors also discuss the challenges that need to be addressed before AI can be widely adopted and developing XAI methods that are effective and efficient in the ATM context.

One of the key challenges is the need for XAI systems to understand the end-user and adapt to their requirements. This involves the ability to interact with the user and provide information that is not only about the internal state of the AI but also caters to the user's needs. Additionally, the format of explanations, such as numeric, rules, textual, and visual explanations, needs to be carefully considered to make the explanation more understandable and user-friendly, which is crucial for effective XAI in ATM.

Furthermore, the interchangeability of terms such as interpretability, transparency, and explainability in the literature poses a hindrance to developing a solid understanding of explainability in AI, which is essential for effective XAI methods in the ATM context. Additionally, the lack of a general guide presenting how to resolve problems in the ATM domain using AI and XAI is a significant limitation that hinders the development of effective and efficient XAI methods in ATM.

Finally, it is important to gain the trust of end users in AI systems. End users need to be confident that AI systems are safe and reliable before they are willing to use them. It is important to develop methods to explain AI systems to end users in a way that builds trust, maintains situational awareness but which doesn’t compromise workload.

5.3 Human Machine Teaming

Human-AI Collaboration/Teaming can be defined as the cooperative and coordinated interaction between humans and artificial intelligence systems to achieve a shared goal or task. It involves the integration of human and AI capabilities, where humans and AI systems work together as interdependent teammates to leverage their respective strengths and expertise [52]. This collaboration can take various forms, such as humans delegating tasks to AI systems, AI systems providing recommendations or assistance to humans, or humans and AI systems jointly making decisions [53].

The goal of human-AI collaboration is to enhance decision-making, problem-solving, and creative processes by combining the strengths of humans and AI systems. It aims at leveraging the computational power, data processing capabilities, and pattern recognition abilities of AI systems, along with the human abilities of critical thinking, creativity, and contextual understanding [54]. Human-AI teaming is being explored in various applications across diverse domains: In military operations, AI is integrated into autonomous systems for target recognition, navigation, and situational awareness, while humans provide strategic guidance and ethical oversight [55], [56]. Moreover, in healthcare, AI assists in medical diagnosis, treatment planning, and patient monitoring, while humans provide expertise and empathy [57], [58]. Additionally, in Air Traffic Management (ATM) [59], [60], where AI has the potential to enhance operational efficiency, careful consideration of factors such as team composition, communication, trust, and the impact of AI on human design teams is essential to ensuring effective human-AI collaboration in ATM. In general, Human-AI teaming is particularly valuable in complex and data-intensive domains, where AI systems can analyse large amounts of data and provide insights, while humans can provide domain expertise, interpret the results, and make informed decisions [53].

5.3.1 Challenges in Human Machine Teaming

© –2023– SESAR 3 JU

While human-AI teaming holds immense promise, it also presents several challenges, as summarised in Table 2. These challenges include:

Table 2: Challenges of Human-AI Collaboration.

Challenges	Category	Description	Source
Design and Interaction	Technical	Designing effective and user-friendly human-AI interactions is difficult due to the uncertainty surrounding AI's capabilities and the complexity of AI outputs.	[61]
Social and Behavioural	Social	Social and behavioural factors, such as trust and confidence in AI systems, can influence the adoption of human-AI collaboration.	[57], [62]
Trust and Confidence	Technical	Establishing trust and confidence in AI systems is crucial for effective human-AI collaboration.	[52], [63]
Ethical and Societal	Ethical	Human-AI collaboration raises ethical and societal concerns, such as bias and discrimination.	[64], [65]
Integration and Coordination	Technical	Integrating and coordinating human and AI capabilities can be challenging due to the need to optimise AI for teamwork and manage human-machine interfaces.	[66], [67]

Design and Interaction Challenges: The first set of challenges is associated with the design and interaction of human-AI systems. Yang et al., in [61] highlight the difficulties in designing valuable human-AI interactions. They propose two main sources of challenge: uncertainty surrounding AI's capabilities and AI's output complexity. AI capabilities can be difficult to predict, and AI systems can produce a wide range of outputs, from simple to adaptive and complex. These challenges can make it difficult for designers to create human-AI interactions that are both effective and user-friendly. While in another study, [68] emphasises the challenges in designing creative AI partners (i.e., co-creative systems for producing creative artefacts, ideas and performances) for Human-AI interaction. These challenges stem from the complexity of human-AI interaction and the need to establish effective communication and collaboration between humans and machines [69].

Social and Behavioural Challenges: Secondly, there are challenges related to the social and behavioural aspects of human-AI collaboration. [57] explores the challenges in healthcare, where social and behavioural factors influence the adoption of human-AI collaboration. Similarly, in [62] the authors highlight the challenges in conducting empathic conversations due to the struggle of AI systems to understand complex human emotions.

Trust and Confidence Challenges: Thirdly, the work in [52] highlight the importance of trust in human-AI collaboration. They suggest that perceived rapport, perceived enjoyment, peer influence, facilitating conditions, and self-efficacy positively affect trust in AI teammates. This study provides insights into the factors that influence trust in human-AI collaboration and can be useful for designing effective

© –2023– SESAR 3 JU

human-AI collaborative systems. A study by [63] discusses the impact of AI advice on human confidence in decision-making. Their findings indicate that the influence of AI advice on human confidence is contingent upon the accuracy of the AI advice. Accurate AI advice supports human confidence in decision-making, while inaccurate AI advice decreases it. The study also revealed that perceived AI advice quality mediates the impact of AI advice on human confidence. These challenges arise from the need to establish trust and confidence in the capabilities and reliability of AI systems.

Ethical and Societal Challenges: The fourth set of challenges are related to the ethical and societal implications of human-AI collaboration. [64] identifies ethical issues in AI partners in human-AI co-creation, while [65] scrutinize the performance and bias in human-AI teamwork in hiring.

Integration and Coordination Challenges: Lastly, there are challenges related to the integration and coordination of human and AI capabilities. The researchers in [66] investigate the optimisation of AI for teamwork, while [67] discusses the collaboration between humans and machines in diverse contexts.

5.3.2 Human AI-Teaming in RDTs

The state-of-the-art human-AI teaming for RDTs involves several key aspects: the integration of advanced automated systems, fail-safe mechanisms, AI technologies, human factors considerations, and the careful design and evaluation of controller-friendly assistance systems.

Firstly, using advanced automated systems requires human operators to efficiently monitor multiple displays with distributed attention and intervene when necessary [70]. This interaction, between operators and machines can be revealed through the path of visual attention, which reflects the cognitive process of human-computer interaction [71]. Additionally, the implementation of fail-safe systems and error control strategies is crucial to reduce the frequency of errors, especially in situations with serious consequences [72].

Furthermore, integrating AI technologies, such as speech recognition support, can significantly enhance the capabilities of air traffic controllers in multiple remote tower environments [19]. For instance, the use of assistant-based speech recognition support (ABSR) can provide real-time assistance to controllers by highlighting recognized callsigns, thereby improving operational efficiency. Moreover, the results of human-in-the-loop experiments have demonstrated that remote tower operations can lead to improvements in communications and departure rates without increasing perceived workload, effort, safety, and situation awareness [8].

In terms of human factors, it is essential to consider the cognitive processing and workload of air traffic controllers in next-generation ATC tower team operations [73]. The SHELL model, can be used to provide a framework for understanding the human factors interfaces in remote tower operations [73]. Additionally, the use of eye movement analysis, including saccades and fixations, can offer insights into the cognitive processes and selective attention of air traffic controllers [74].

Moreover, the design and evaluation of a controller-friendly assistance system, along with the careful selection of test participants, are critical for obtaining relevant and objective results in developing remote tower solutions [72]. Furthermore, considering user training and habituation to virtual environments can help reduce cybersickness and improve user experience [75].

© –2023– SESAR 3 JU

5.4 Summary of TIS

For the TRUSTY project, this SotA report integrates insights from our previous ARTIMATION project and the evolving landscape of Robust and interpretable AI, HCXAI and Human-AI Teaming (HAIT), specifically in the context of Taxiway and Runway monitoring tasks within RDT. Table 2 shows a common list of categories and methods concerning trustworthy AI.

The TRUSTY project can greatly benefit from using robust and interpretable AI, especially for tasks that require reliability and increased insight into AI decision-making rationale. MML will be important for making this rationale clear and fair. [37] pointed out that MML can be used for multimedia event detection, which includes video summarization, action classification and multimedia content analysis. [39] introduced a variety of advanced tools like Multimodal Knowledge Graphs and Generative Adversarial Networks (GANs) that can help with visual answering questions, filling in knowledge gaps, and describing videos and images, which are useful for the project's goals.

Drawing from the principles of HCXAI, TRUSTY prioritises creating AI explanations that are understandable, actionable, and trustworthy. It emphasises the need for AI systems to meet the needs of both AI experts and non-experts, offering transparent and intuitive explanations that enhance user trust and decision-making. Moreover, the project will explore the HCXAI approaches. TRUSTY recognises the complexity of evaluating XAI techniques and will focus on specific methods to assess user perceptions of AI explainability. Human-centric evaluation methods will centre around trust, explanation of usefulness and satisfaction, understandability, and human performance. The project will also address the challenges of ensuring explainability in increasingly complex AI systems, focusing on the balance between accuracy and transparency.

The project will explore the cooperative interaction between humans and AI in achieving shared goals. It will explore the RDT domain, emphasizing the integration of human critical thinking and the computational power of AI. Challenges in design, interaction, social and behavioural aspects, trust, ethics, and integration will be explored. The project aims to enhance decision-making and problem-solving by leveraging the strengths of both humans and AI.

In the context of RDT, TRUSTY will apply HCXAI to improve operational efficiency and decision-making. It will draw insights from related fields like smart cities and industrial asset management, focusing on transparency, accountability, and effective use of XAI techniques.

Building on the findings of the ARTIMATION project's findings in the field of AI and XAI for Air Traffic Management (ATM), TRUSTY will incorporate lessons learned in developing transparent and understandable AI systems.

In conclusion, the TRUSTY project aims to advance robust and interpretable AI, HCXAI and HAIT by integrating key learnings from the ARTIMATION project and recent developments in the field. It seeks to foster trust, understanding, and effective collaboration between humans and AI systems, focusing on remote digital towers in the context of ATMs. This SotA report lays the foundation for the technical roadmap and future directions of the TRUSTY project.

Table 3: List of categories and methods.

Category	Method	Application	Paper

© –2023– SESAR 3 JU

Image and vision	Convolutional Neural Networks (CNNs)	Image captioning, visual question-answering, and multimedia event detection.	[37]
		Visual Question Answering (VQA), Visual Commonsense Reasoning, Visual Dialogue, Phrase Grounding	[38]
	Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs)	Image classification, object detection, image captioning, and visual question answering	[39]
Text	Recurrent Neural Networks (RNNs) for NLP	Attention mechanisms in generating image captions, deep learning for learning image-text embeddings.	[37]
	NLP	Visual Question Answering (VQA), Visual Commonsense Reasoning, Visual Dialogue, Phrase Grounding	[38]
	Recurrent Neural Networks (RNNs) and Transformer models, Word embeddings (e.g., Word2Vec, GloVe)	Sentiment analysis, machine translation, text summarization, and document classification	[39]
MML	Multimodal Knowledge Graphs, Multimodal Commonsense Reasoning, Attention Mechanisms, Graph-based Methods, generative adversarial networks (GANs), variational autoencoders	Event detection	[37]

© –2023– SESAR 3 JU
EUROPEAN PARTNERSHIP
Co-funded by the European Union

	(VAEs), Matrix Factorization		
		Visual question answering, knowledge base completion, and image captioning, recommendation systems and data clustering	[39]
Fairness	Fairness measuring methods: statistical parity, disparate impact, equalized odds, calibration, counterfactual fairness	Recidivism Prediction, Credit Scoring, Employment and Hiring and many others	[41]

© –2023– SESAR 3 JU
EUROPEAN PARTNERSHIP
Co-funded by the European Union

6 Human Factor and Cognitive Assessment

As AI technology becomes an increasingly important part of our everyday working life, human work is increasingly influenced by AI, and thus there is a growing need to effectively integrate, collaborate and connect with it [76]. AI technologies refer to intelligent systems that perform human cognitive functions such as learning, interaction, problem solving and decision making, and therefore can be used with the same flexibility as human workers [77]. As technology advances, AI can be integrated directly into team processes alongside other artificial and human agents or perform roles that assist humans in the same way as team members. This interaction is defined as HAIT [78]. The different but complementary capabilities of the human-AI team help to work together effectively to achieve complex goals while ensuring people's well-being, motivation, and productivity. Other synergies that arise when operators work together with AI contribute to strategic decision-making [79], the development of individual competences and thus long-term employee motivation [80]. Employee acceptance and positive attitudes towards working with AI increase when AI is considered as a team member. HAIT therefore offers an opportunity to create attractive and sustainable workplaces by utilizing human capabilities, providing learning and mutual support.

However, these benefits are not obvious when humans are in a team with AI systems. The National Academies of Sciences, Engineering, and Medicine [81] identifies four conditions for humans and AI teams to benefit from synergy.

(1) Humans must be able to understand and predict the behaviour of the intelligent agents employed.
(2) People must be able to establish appropriate trustiness to properly use artificial intelligence systems.
(3) People must be able to make accurate decisions when using the results of employed systems.
(4) People must be able to control and manage the system appropriately.

These conditions show that successful teamwork depends not only on technical aspects (e.g., design of the AI system) but also on human aspects (e.g., mental workload and stress induced during the interaction, trust in the system) that may induce poor interaction/teamwork (e.g., form of collaboration).

With a particular regard to the human aspects involved in HAIT (i.e., human factors) to take into account during the interaction with AI, neurophysiological measurement, based on the recording of operators' neurophysiological signals, (e.g., EEG, ECG, EDA), showed so far clear advantages with respect to other methodologies, such as subjective or performance measures [82]. Firstly, neurophysiological measures could be obtained continuously and online. Secondly, compared with subjective and performance measures, the neurophysiological ones may be recorded continuously without using overt responses (i.e., additional tasks) and may provide a direct measure of the mental (covert) activities of the operator during the interaction with the AI. Also, neurophysiological measures have higher resolution than subjective and performance measures [83]. Last, but not least, the big advantage of this kind of measures lies in the possibility to provide information coming from the operator (i.e., mental, and emotional states) directly in input to the AI, to make it more empathetic, and so inducing an enhancement of the HAIT itself.

© –2023– SESAR 3 JU

Byrne and Parasuraman in [82] assessed that the advantage of applying neurophysiological measures in triggering AI was very clear, but the "*effective application of psychophysiology in the regulatory role may require years of effort and considerable maturation in technology*." Today, 30 years later, this "effective application" could become a reality thanks to advances in brain-computer interface (BCI) research.

Briefly, a BCI is defined as "a system that measures Central Nervous System (CNS) activity and converts it into artificial output that replaces, restores, enhances or improves natural CNS output and thereby changes the ongoing interactions between the CNS and its external or internal environment" [84]. Such definition summarizes the progresses of the scientific community in this field during the last decades, since at the moment the possibility of using the BCI systems outside the laboratories [85], by developing applications in everyday life is not just a theory but something very close to real applications [86]. This technology has been defined passive Brain-Computer Interface (pBCI). In pBCI technologies, the system recognizes the spontaneous brain activity of the user related to the considered mental state (e.g., emotional state, workload, attention levels), and uses such information to improve and modulate the interaction between the operator and the system (i.e., AI) itself. Thus, in the context of AI, the pBCIs perfectly match the needs of the system in terms of HAIT [87], [88].

To enhance the trustworthiness of the remote tower operators in the AI, and so to maximize the HAIT effectiveness, the TRUSTY project will employ the pBCI concept. It is possible to derive real-time information from the ongoing brain activity of the operator, by using signals coming from the body (e.g., brain signals), while he/she is doing his/her operational activity (i.e. remote tower operations). From such signals, it is possible to evaluate specific metrics (i.e., neurometrics) that correlate with a variation of mental and emotional states of the user, such as workload, stress, vigilance, acceptance, and this information can be used online, to modify the behaviour of the interface (and AI) that the operator is interacting with. This system will be able to put the operator in the loop, so that the AI model can adapt its behaviour, by considering the actual mental or emotional state of the user, with an increased trustworthiness in the AI, resulting in a powered Human-AI-Teaming. For example, the level of explainability of the XAI could be adapted to the actual state of the user (e.g., low workload and stress, could correspond to a high level of explanation or vice versa). In addition, the AI itself can use as input features the information coming in real time from the user states, together with all the other parameters coming from the HMI (e.g., the traffic).

A further theme concerning the cognitive assessment of human-machine teaming, and particularly HAIT, is the possibility to assess through neurophysiological measures the unconscious attitude toward the employment of AI. In fact, despite the human-AI relationship and interaction is welcome in some areas [89], recent evidence shows both explicit and implicit bias towards AI [90], [91], [92]. For instance, art, one of the most evolved and complex amongst human activities - just as science is - and difficult to be implemented by an artificial agent, represents a good model for studying Human-AI interaction [93]. With respect to this, the first study investigating the negative bias toward AI use in neuroesthetics showed that when human and AI products are compared, emotional arousal measured through neurophysiological measures (i.e., electrodermal activity) increases and a negative bias toward AI emerges in declarative ratings [94]. Such evidence suggests EDA increase as a psychophysiological activation plausibly induced by the recruitment of implicit comparison mechanisms, supporting the sensitivity of EDA in detecting unconscious reaction during ambiguous choices and categorization [95]. Such autonomic signal-derived reactivity in response to items, possibly objects of biased prejudice, has been also observed when comparing foreign and local products [96].

© –2023– SESAR 3 JU
EUROPEAN PARTNERSHIP

Co-funded by the European Union

In the just mentioned article, it is also suggested the sensitivity of EEG frontal alpha asymmetry index and frontal theta index in detecting the attitude toward the adoption of familiar and unfamiliar products, that can be obviously extended to the assessment of the propensity, that is HAIT, in human-machine team actions.

© –2023– SESAR 3 JU

EUROPEAN PARTNERSHIP

Co-funded by the European Union

7 Conclusions

This report represents the delivery D 3.1- Report on Definition, Specification and SoTA in RDTs under work package (WP) WP3. The object of this WP is to define the specifications regarding trustworthy AI solutions in the operational activities in RDT, especially in runway and taxiway monitoring.

The report presents a general background of related work in the RDTs, defines some use-case scenarios, and delves into the various aspects of trustworthy AI and machine learning. It also provides an overview of human-machine teaming and the utilization of cognitive assessment to understand the effects of HCAI in RDTs' operations.

To exemplify, an AI system, like a Chatbot or a system that answers questions or makes recommendations, can use Large Language Models (LLM) technology, using its own data to help with a specific Remote Digital Tower task.

The AI system can provide the following capabilities:

(1) **ChatBot:** It can help in communication between controllers and pilots, making it more efficient and accurate.
(2) **Question Answering (QA):** The AI system can answer questions from controllers or pilots about weather, flight schedules, or other important information.
(3) **Recommendation Systems:** It can provide recommendations based on real-time data, helping controllers make decisions about landing, take-off, and airport operations.

Even if LLM can present challenges, including ensuring that the AI system is reliable, secure from cyber threats, and free from biases. It is believed that, overall, using LLM technology in an RDT can improve aviation operations and safety.

The Trusty project will explore the challenges, considering the scenarios, based on RDTs operations, defined in this report, and develop MML models incorporating explainability, fairness, auditability and accountability of AI system. To make the AI system transparent, one important aspect the Trusty project considers is human-machine teaming so that the AI system is trustworthy by design.

© –2023– SESAR 3 JU
EUROPEAN PARTNERSHIP
Co-funded by the European Union

8 References

[1] Y. Rong *et al.*, 'Towards Human-centered Explainable AI: A Survey of User Studies for Model Explanations'. arXiv, Sep. 05, 2023. Accessed: Oct. 01, 2023. [Online]. Available: http://arxiv.org/abs/2210.11584

[2] L. Bianco and M. Bielli, 'Air traffic management: Optimization models and algorithms', *Journal of Advanced Transportation*, vol. 26, no. 2, pp. 131–167, 1992, doi: 10.1002/atr.5670260205.

[3] S. Ratcliffe, 'Automation in Air Traffic Management', *The Journal of Navigation*, vol. 38, no. 3, pp. 405–412, Sep. 1985, doi: 10.1017/S0373463300032768.

[4] B. Sridhar and G. B. Chatterji, 'Computationally efficient conflict detection methods for air traffic management', in *Proceedings of the 1997 American Control Conference (Cat. No.97CH36041)*, Jun. 1997, pp. 1126–1130 vol.2. doi: 10.1109/ACC.1997.609708.

[5] J. DeArmon, 'Aggregation effects in air traffic arrival flows', *Journal of Guidance, Control, and Dynamics*, vol. 18, no. 3, pp. 579–583, May 1995, doi: 10.2514/3.21426.

[6] T. Standfuss and J. Rosenow, 'Applicability of Current Complexity Metrics in ATM Performance Benchmarking and Potential Benefits of Considering Weather Conditions', in *2020 AIAA/IEEE 39th Digital Avionics Systems Conference (DASC)*, Oct. 2020, pp. 1–9. doi: 10.1109/DASC50938.2020.9256719.

[7] P. Kearney and W.-C. Li, 'Multiple remote tower for Single European Sky: The evolution from initial operational concept to regulatory approved implementation', *Transportation Research Part A: Policy and Practice*, vol. 116, pp. 15–30, Oct. 2018, doi: 10.1016/j.tra.2018.06.005.

[8] W.-C. Li, P. Kearney, G. Braithwaite, and J. J. H. Lin, 'How much is too much on monitoring tasks? Visual scan patterns of single air traffic controller performing multiple remote tower operations', *International Journal of Industrial Ergonomics*, vol. 67, pp. 135–144, Sep. 2018, doi: 10.1016/j.ergon.2018.05.005.

[9] M. Cordeil, T. Dwyer, and C. Hurter, 'Immersive solutions for future Air Traffic Control and Management', in *Proceedings of the 2016 ACM Companion on Interactive Surfaces and Spaces*, in ISS '16 Companion. New York, NY, USA: Association for Computing Machinery, Nov. 2016, pp. 25–31. doi: 10.1145/3009939.3009944.

[10] M. Reynal *et al.*, 'Investigating Multimodal Augmentations Contribution to Remote Control Tower Contexts for Air Traffic Management', presented at the HUCAPP 2019, 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Scitepress, 2019, p. pp 50. doi: 10.5220/0007400300500061.

[11] D. Eier, H. Huber, and W. Kampichler, 'Advanced ground surveillance for remote tower', in *2008 Integrated Communications, Navigation and Surveillance Conference*, Bethesda, MD, USA: IEEE, May 2008, pp. 1–9. doi: 10.1109/ICNSURV.2008.4559159.

© –2023– SESAR 3 JU

[12] M. Schmidt *et al.*, 'Remote airport traffic control center with augmented vision video panorama', in *2009 IEEE/AIAA 28th Digital Avionics Systems Conference*, Orlando, FL: IEEE, Oct. 2009, p. 4.E.2-1-4.E.2-15. doi: 10.1109/DASC.2009.5347479.

[13] M. C. Moruzzi, S. Santhosh, M. Corsi, S. Bagassi, and F. De Crescenzio, 'Design and implementation of eye tracking application for generation of augmented reality content on spatial see through display of remote and virtual control tower (RVT)', *Int J Interact Des Manuf*, vol. 17, no. 4, pp. 1859–1867, Aug. 2023, doi: 10.1007/s12008-023-01288-7.

[14] N. Fürstenau, M. Friedrich, M. Mittendorf, M. Schmidt, and M. Rudolph, 'Discriminability of Flight Maneuvers and Risk of False Decisions Derived from Dual Choice Decision Errors in a Videopanorama-Based Remote Tower Work Position', in *Engineering Psychology and Cognitive Ergonomics. Applications and Services*, D. Harris, Ed., in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2013, pp. 105–114. doi: 10.1007/978-3-642-39354-9_12.

[15] C. Westin, S. Boonsong, B. Josefsson, and J. Lundberg, 'Building Trust in Autonomous System Competence - the DiTA Digital Tower Assistant for Multiple Remote Towers, an Early Concept Evaluation', presented at the 39th AIAA/IEEE Digital Avionics Systems Conference (DASC), ELECTR NETWORK, oct 11-16, 2020, IEEE, 2020. Accessed: Feb. 07, 2024. [Online]. Available: https://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-175950

[16] V.-P. Thai, W. Zhong, T. Pham, S. Alam, and V. Duong, 'Detection, Tracking and Classification of Aircraft and Drones in Digital Towers Using Machine Learning on Motion Patterns', in *2019 Integrated Communications, Navigation and Surveillance Conference (ICNS)*, Apr. 2019, pp. 1–8. doi: 10.1109/ICNSURV.2019.8735240.

[17] L. Meyer, M. Peukert, B. Josefsson, and J. Lundberg, 'Validation of an empiric method for safety assessment of multi remote tower', presented at the 13th USA/Europe Air Traffic Management Research and Development Seminar 2019, ATM 2019, Vienna, 17 June 2019 - 21 June 2019, EUROCONTROL, 2019. Accessed: Feb. 07, 2024. [Online]. Available: https://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-179114

[18] M. Hagl *et al.*, 'Augmented Reality in a Remote Tower Environment Based on VS/IR Fusion and Optical Tracking', in *Engineering Psychology and Cognitive Ergonomics*, D. Harris, Ed., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018, pp. 558–571. doi: 10.1007/978-3-319-91122-9_45.

[19] O. Ohneiser *et al.*, 'Assistant Based Speech Recognition Support for Air Traffic Controllers in a Multiple Remote Tower Environment', *Aerospace*, vol. 10, no. 6, p. 560, Jun. 2023, doi: 10.3390/aerospace10060560.

[20] J. Zhang, X. Tian, J. Pan, Z. Chen, and X. Zou, 'A Field Study on Safety Performance of Apron Controllers at a Large-Scale Airport Based on Digital Tower', *International Journal of Environmental Research and Public Health*, vol. 19, no. 3, Art. no. 3, Jan. 2022, doi: 10.3390/ijerph19031623.

[21] W. Pan, Y. Luo, S. Han, and L. Zhong, 'Summary of Reliability Analysis Methods of Remote Tower System', *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 693, no. 1, p. 012062, Mar. 2021, doi: 10.1088/1755-1315/693/1/012062.

© –2023– SESAR 3 JU

EUROPEAN PARTNERSHIP

Co-funded by the European Union

[22] A. Disdier, D. Masson, T. Brethomé, M. Jankovic, and G. A. Boy, 'Toward a Human Systems Integration approach to the design and operation of a remote and virtual air traffic control center', in *24th International Conference on Engineering Design (ICED23)*, Bordeaux, France: The Design Society, Jul. 2023. Accessed: Feb. 07, 2024. [Online]. Available: https://hal.science/hal-04064881

[23] D. Gunning and D. Aha, 'DARPA's Explainable Artificial Intelligence (XAI) Program', *AI Magazine*, vol. 40, no. 2, Art. no. 2, Jun. 2019, doi: 10.1609/aimag.v40i2.2850.

[24] M. Ridley, 'Explainable Artificial Intelligence (XAI): Adoption and Advocacy', *Information Technology and Libraries*, vol. 41, no. 2, Art. no. 2, Jun. 2022, doi: 10.6017/ital.v41i2.14683.

[25] F. Kaptein, J. Broekens, K. Hindriks, and M. Neerincx, 'Personalised self-explanation by robots: The role of goals versus beliefs in robot-action explanation for children and adults', in *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Aug. 2017, pp. 676–682. doi: 10.1109/ROMAN.2017.8172376.

[26] A. Barredo Arrieta *et al.*, 'Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI', *Information Fusion*, vol. 58, pp. 82–115, Jun. 2020, doi: 10.1016/j.inffus.2019.12.012.

[27] R. Confalonieri, L. Coba, B. Wagner, and T. R. Besold, 'A historical perspective of explainable Artificial Intelligence', *WIREs Data Mining and Knowledge Discovery*, vol. 11, no. 1, p. e1391, 2021, doi: 10.1002/widm.1391.

[28] V. Adarsh and G. R. Gangadharan, 'Applying Explainable Artificial Intelligence Models for Understanding Depression Among IT Workers', *IT Professional*, vol. 24, no. 5, pp. 25–29, Sep. 2022, doi: 10.1109/MITP.2022.3209803.

[29] F. Cruz, C. Young, R. Dazeley, and P. Vamplew, 'Evaluating Human-like Explanations for Robot Actions in Reinforcement Learning Scenarios'. arXiv, Jul. 07, 2022. doi: 10.48550/arXiv.2207.03214.

[30] T. Miller, 'Explanation in artificial intelligence: Insights from the social sciences', *Artificial Intelligence*, vol. 267, pp. 1–38, Feb. 2019, doi: 10.1016/j.artint.2018.07.007.

[31] A. Degas *et al.*, 'A Survey on Artificial Intelligence (AI) and eXplainable AI in Air Traffic Management: Current Trends and Development with Future Research Trajectory', *Applied Sciences*, vol. 12, no. 3, Art. no. 3, Jan. 2022, doi: 10.3390/app12031295.

[32] 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.

[33] 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.

[34] 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.

© –2023– SESAR 3 JU

[35] 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.

[36] 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

[37] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, 'Multimodal Machine Learning: A Survey and Taxonomy'. arXiv, Aug. 01, 2017. Accessed: Oct. 30, 2023. [Online]. Available: http://arxiv.org/abs/1705.09406

[38] L. Parcalabescu, N. Trost, and A. Frank, 'What is Multimodality?' arXiv, Jun. 10, 2021. Accessed: Oct. 30, 2023. [Online]. Available: http://arxiv.org/abs/2103.06304

[39] P. P. Liang, A. Zadeh, and L.-P. Morency, 'Foundations and Trends in Multimodal Machine Learning: Principles, Challenges, and Open Questions'. arXiv, Feb. 20, 2023. Accessed: Oct. 30, 2023. [Online]. Available: http://arxiv.org/abs/2209.03430

[40] D. Hendrycks *et al.*, 'The Many Faces of Robustness: A Critical Analysis of Out-of-Distribution Generalization'. arXiv, Jul. 24, 2021. Accessed: Sep. 29, 2023. [Online]. Available: http://arxiv.org/abs/2006.16241

[41] S. Caton and C. Haas, 'Fairness in Machine Learning: A Survey'. arXiv, Oct. 04, 2020. Accessed: Sep. 29, 2023. [Online]. Available: http://arxiv.org/abs/2010.04053

[42] J. A. Pérez-Castán, L. Pérez Sanz, M. Fernández-Castellano, T. Radišić, K. Samardžić, and I. Tukarić, 'Learning Assurance Analysis for Further Certification Process of Machine Learning Techniques: Case-Study Air Traffic Conflict Detection Predictor', *Sensors*, vol. 22, no. 19, p. 7680, Oct. 2022, doi: 10.3390/s22197680.

[43] S. Wandelt, X. Sun, and A. Zhang, 'AI-driven assistants for education and research? A case study on ChatGPT for air transport management', *Journal of Air Transport Management*, vol. 113, p. 102483, Oct. 2023, doi: 10.1016/j.jairtraman.2023.102483.

[44] A. Das and P. Rad, 'Opportunities and Challenges in Explainable Artificial Intelligence (XAI): A Survey'. arXiv, Jun. 22, 2020. doi: 10.48550/arXiv.2006.11371.

[45] Q. V. Liao and K. R. Varshney, 'Human-Centered Explainable AI (XAI): From Algorithms to User Experiences'. arXiv, Apr. 19, 2022. Accessed: Oct. 01, 2023. [Online]. Available: http://arxiv.org/abs/2110.10790

[46] U. Ehsan and M. O. Riedl, 'Human-Centered Explainable AI: Towards a Reflective Sociotechnical Approach', in *HCI International 2020 - Late Breaking Papers: Multimodality and Intelligence*, vol. 12424, C. Stephanidis, M. Kurosu, H. Degen, and L. Reinerman-Jones, Eds., in Lecture Notes in Computer Science, vol. 12424. , Cham: Springer International Publishing, 2020, pp. 449–466. doi: 10.1007/978-3-030-60117-1_33.

© –2023– SESAR 3 JU

[47] B. Shneiderman, 'Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy', *International Journal of Human–Computer Interaction*, vol. 36, no. 6, pp. 495–504, Apr. 2020, doi: 10.1080/10447318.2020.1741118.

[48] P. Lopes, E. Silva, C. Braga, T. Oliveira, and L. Rosado, 'XAI Systems Evaluation: A Review of Human and Computer-Centred Methods', *Applied Sciences*, vol. 12, no. 19, p. 9423, Sep. 2022, doi: 10.3390/app12199423.

[49] E. Small, Y. Xuan, D. Hettiachchi, and K. Sokol, 'Helpful, Misleading or Confusing: How Humans Perceive Fundamental Building Blocks of Artificial Intelligence Explanations'. arXiv, Apr. 15, 2023. Accessed: Feb. 05, 2024. [Online]. Available: http://arxiv.org/abs/2303.00934

[50] A. R. Javed, W. Ahmed, S. Pandya, P. K. R. Maddikunta, M. Alazab, and T. R. Gadekallu, 'A Survey of Explainable Artificial Intelligence for Smart Cities', *Electronics*, vol. 12, no. 4, Art. no. 4, Jan. 2023, doi: 10.3390/electronics12041020.

[51] A. K. M. Nor, S. R. Pedapati, M. Muhammad, and V. Leiva, 'Overview of Explainable Artificial Intelligence for Prognostic and Health Management of Industrial Assets Based on Preferred Reporting Items for Systematic Reviews and Meta-Analyses', *Sensors*, vol. 21, no. 23, Art. no. 23, Jan. 2021, doi: 10.3390/s21238020.

[52] K. Hou, T. Hou, and L. Cai, 'Exploring Trust in Human–AI Collaboration in the Context of Multiplayer Online Games', *Systems*, vol. 11, no. 5, Art. no. 5, May 2023, doi: 10.3390/systems11050217.

[53] V. Lai, S. Carton, R. Bhatnagar, Q. V. Liao, Y. Zhang, and C. Tan, 'Human-AI Collaboration via Conditional Delegation: A Case Study of Content Moderation', in *CHI Conference on Human Factors in Computing Systems*, New Orleans LA USA: ACM, Apr. 2022, pp. 1–18. doi: 10.1145/3491102.3501999.

[54] D. Siemon, 'Elaborating Team Roles for Artificial Intelligence-based Teammates in Human-AI Collaboration', *Group Decis Negot*, vol. 31, no. 5, pp. 871–912, Oct. 2022, doi: 10.1007/s10726-022-09792-z.

[55] F. L. Oswald *et al.*, 'The National Academies Board on Human-Systems Integration (BOHSI) Panel: Human-AI Teaming: Research Frontiers', *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 66, no. 1, pp. 130–134, Sep. 2022, doi: 10.1177/1071181322661007.

[56] S. Sawant, C. Brady, R. Mallick, N. McNeese, K. Chalil Madathil, and J. Bertrand, 'Human-AI teams in complex military operations: Soldiers' perception of intelligent AI agents as teammates in human-AI teams', *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, p. 21695067231192423, Oct. 2023, doi: 10.1177/21695067231192423.

[57] Y. Lai, A. Kankanhalli, and D. Ong, 'Human-AI Collaboration in Healthcare: A Review and Research Agenda', *Hawaii International Conference on System Sciences 2021 (HICSS-54)*, Jan. 2021, [Online]. Available: https://aisel.aisnet.org/hicss-54/cl/machines_as_teammates/5

[58] K. E. Henry *et al.*, 'Human–machine teaming is key to AI adoption: clinicians' experiences with a deployed machine learning system', *npj Digit. Med.*, vol. 5, no. 1, Art. no. 1, Jul. 2022, doi: 10.1038/s41746-022-00597-7.

© –2023– SESAR 3 JU

[59] S. Malakis *et al.*, 'A Framework for Supporting Adaptive Human-AI Teaming in Air Traffic Control', in *Engineering Psychology and Cognitive Ergonomics*, D. Harris and W.-C. Li, Eds., in Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, 2023, pp. 320–330. doi: 10.1007/978-3-031-35389-5_22.

[60] N. Minaskan, C. Alban-Dromoy, A. Pagani, J.-M. Andre, and D. Stricker, 'Human Intelligent Machine Teaming in Single Pilot Operation: A Case Study', in *Augmented Cognition*, D. D. Schmorrow and C. M. Fidopiastis, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2022, pp. 348–360. doi: 10.1007/978-3-031-05457-0_27.

[61] Q. Yang, A. Steinfeld, C. Rosé, and J. Zimmerman, 'Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design', in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, in CHI '20. New York, NY, USA: Association for Computing Machinery, Apr. 2020, pp. 1–13. doi: 10.1145/3313831.3376301.

[62] A. Sharma, I. W. Lin, A. S. Miner, D. C. Atkins, and T. Althoff, 'Human-AI Collaboration Enables More Empathic Conversations in Text-based Peer-to-Peer Mental Health Support'. arXiv, Mar. 28, 2022. doi: 10.48550/arXiv.2203.15144.

[63] A. Taudien, A. Fügener, A. Gupta, and W. Ketter, 'The Effect of AI Advice on Human Confidence in Decision-Making', presented at the Hawaii International Conference on System Sciences, 2022. doi: 10.24251/HICSS.2022.029.

[64] J. Rezwana and M. L. Maher, 'Identifying Ethical Issues in AI Partners in Human-AI Co-Creation'. arXiv, Apr. 15, 2022. doi: 10.48550/arXiv.2204.07644.

[65] A. Peng, B. Nushi, E. Kiciman, K. Inkpen, and E. Kamar, 'Investigations of Performance and Bias in Human-AI Teamwork in Hiring'. arXiv, Feb. 21, 2022. doi: 10.48550/arXiv.2202.11812.

[66] G. Bansal, B. Nushi, E. Kamar, E. Horvitz, and D. S. Weld, 'Is the Most Accurate AI the Best Teammate? Optimizing AI for Teamwork', *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 13, Art. no. 13, May 2021, doi: 10.1609/aaai.v35i13.17359.

[67] J. Makokha, 'FUTURE AI SYSTEMS: AN AI OUTPERFORMS A HUMAN AS TEAMMATE'. TechRxiv, Jan. 26, 2023. doi: 10.36227/techrxiv.21901407.v1.

[68] J. Rezwana and M. L. Maher, 'Designing Creative AI Partners with COFI: A Framework for Modeling Interaction in Human-AI Co-Creative Systems', *ACM Trans. Comput.-Hum. Interact.*, vol. 30, no. 5, p. 67:1-67:28, Sep. 2023, doi: 10.1145/3519026.

[69] J. Rezwana and M. L. Maher, 'Understanding User Perceptions, Collaborative Experience and User Engagement in Different Human-AI Interaction Designs for Co-Creative Systems', in *Proceedings of the 14th Conference on Creativity and Cognition*, in C&C '22. New York, NY, USA: Association for Computing Machinery, Jun. 2022, pp. 38–48. doi: 10.1145/3527927.3532789.

[70] C. Bruder, H. Eißfeldt, P. Maschke, and C. Hasse, 'A model for future aviation: Operators monitoring appropriately', *Aviation Psychology and Applied Human Factors*, vol. 4, no. 1, pp. 13–22, 2014, doi: 10.1027/2192-0923/a000051.

© –2023– SESAR 3 JU

[71] P. Kearney, W.-C. Li, and J. J. H. Lin, 'The impact of alerting design on air traffic controllers' response to conflict detection and resolution', *International Journal of Industrial Ergonomics*, vol. 56, pp. 51–58, Nov. 2016, doi: 10.1016/j.ergon.2016.09.002.

[72] R. Leitner and A. Oehme, 'Planning Remote Multi-airport Control—Design and Evaluation of a Controller-Friendly Assistance System', in *Virtual and Remote Control Tower*, N. Fürstenau, Ed., Cham: Springer International Publishing, 2022, pp. 489–510. doi: 10.1007/978-3-030-93650-1_20.

[73] M. Miller, S. Holley, B. Mrusek, and L. Weiland, 'Assessing Cognitive Processing and Human Factors Challenges in NextGen Air Traffic Control Tower Team Operations', in *Advances in Human Factors and Systems Interaction*, vol. 1207, I. L. Nunes, Ed., Cham: Springer International Publishing, 2020, pp. 289–295. doi: 10.1007/978-3-030-51369-6_39.

[74] H.-J. Wee, S.-W. Lye, and J.-P. Pinheiro, 'Computing Eye Tracking Metric for a Radar Display Using a Remote Eye Tracker', *DEStech Transactions on Computer Science and Engineering*, vol. 0, no. aice-ncs, Art. no. aice-ncs, 2016, doi: 10.12783/dtcse/aice-ncs2016/5671.

[75] F. Reuschling and J. Jakobi, 'Designing a Low-Cost Remote Tower Solution', in *Virtual and Remote Control Tower*, N. Fürstenau, Ed., Cham: Springer International Publishing, 2022, pp. 543–566. doi: 10.1007/978-3-030-93650-1_22.

[76] N. J. McNeese, M. Demir, N. J. Cooke, and M. She, 'Team Situation Awareness and Conflict: A Study of Human–Machine Teaming', *Journal of Cognitive Engineering and Decision Making*, vol. 15, no. 2–3, pp. 83–96, Sep. 2021, doi: 10.1177/15553434211017354.

[77] D. Dellermann, A. Calma, N. Lipusch, T. Weber, S. Weigel, and P. Ebel, 'The future of human-AI collaboration: a taxonomy of design knowledge for hybrid intelligence systems'. arXiv, May 07, 2021. doi: 10.48550/arXiv.2105.03354.

[78] N. J. McNeese, M. Demir, N. J. Cooke, and C. Myers, 'Teaming With a Synthetic Teammate: Insights into Human-Autonomy Teaming', *Hum Factors*, vol. 60, no. 2, pp. 262–273, Mar. 2018, doi: 10.1177/0018720817743223.

[79] P. Aversa, L. Cabantous, and S. Haefliger, 'When decision support systems fail: Insights for strategic information systems from Formula 1', *The Journal of Strategic Information Systems*, vol. 27, no. 3, pp. 221–236, Sep. 2018, doi: 10.1016/j.jsis.2018.03.002.

[80] A. Kluge, G. Ontrup, V. Langholf, and U. Wilkens, 'Mensch-KI-Teaming: Mensch und Künstliche Intelligenz in der Arbeitswelt von morgen', *Zeitschrift für wirtschaftlichen Fabrikbetrieb*, vol. 116, no. 10, pp. 728–734, Oct. 2021, doi: 10.1515/zwf-2021-0112.

[81] *Human-AI Teaming: State-of-the-Art and Research Needs*. Washington, D.C.: National Academies Press, 2022. doi: 10.17226/26355.

[82] E. A. Byrne and R. Parasuraman, 'Psychophysiology and adaptive automation', *Biological Psychology*, vol. 42, no. 3, pp. 249–268, Feb. 1996, doi: 10.1016/0301-0511(95)05161-9.

[83] G. Di Flumeri *et al.*, 'On the Use of Cognitive Neurometric Indexes in Aeronautic and Air Traffic Management Environments', in *Symbiotic Interaction*, B. Blankertz, G. Jacucci, L. Gamberini, A.

© –2023– SESAR 3 JU

Spagnolli, and J. Freeman, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015, pp. 45–56. doi: 10.1007/978-3-319-24917-9_5.

[84] J. Wolpaw, E. W. Wolpaw, J. Wolpaw, and E. W. Wolpaw, Eds., *Brain-Computer Interfaces: Principles and Practice*. Oxford, New York: Oxford University Press, 2012.

[85] P. Aricò *et al.*, 'GeoSpell: an alternative P300 based speller interface towards no eye gaze required', *International Journal of Bioelectromagnetism www.ijbem.org*, vol. 13, pp. 152–153, Jan. 2011.

[86] N. Sciaraffa *et al.*, 'Evaluation of a New Lightweight EEG Technology for Translational Applications of Passive Brain-Computer Interfaces', *Frontiers in Human Neuroscience*, vol. 16, 2022, Accessed: Feb. 08, 2024. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnhum.2022.901387

[87] T. O. Zander, C. Kothe, S. Welke, and M. Roetting, 'Utilizing Secondary Input from Passive Brain-Computer Interfaces for Enhancing Human-Machine Interaction', in *Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience*, D. D. Schmorrow, I. V. Estabrooke, and M. Grootjen, Eds., in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2009, pp. 759–771. doi: 10.1007/978-3-642-02812-0_86.

[88] D. D. Schmorrow, I. V. Estabrooke, and M. Grootjen, *Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience: 5th International Conference, FAC 2009, Held as Part of HCI International 2009 San Diego, CA, USA, July 19-24, 2009, Proceedings*. Springer Science & Business Media, 2009.

[89] N. Tomašev *et al.*, 'AI for social good: unlocking the opportunity for positive impact', *Nat Commun*, vol. 11, no. 1, Art. no. 1, May 2020, doi: 10.1038/s41467-020-15871-z.

[90] V. Fietta, F. Zecchinato, B. D. Stasi, M. Polato, and M. Monaro, 'Dissociation Between Users' Explicit and Implicit Attitudes Toward Artificial Intelligence: An Experimental Study', *IEEE Transactions on Human-Machine Systems*, vol. 52, no. 3, pp. 481–489, Jun. 2022, doi: 10.1109/THMS.2021.3125280.

[91] Y. Liang and S. A. Lee, 'Fear of Autonomous Robots and Artificial Intelligence: Evidence from National Representative Data with Probability Sampling', *Int J of Soc Robotics*, vol. 9, no. 3, pp. 379–384, Jun. 2017, doi: 10.1007/s12369-017-0401-3.

[92] L. Sartori and G. Bocca, 'Minding the gap(s): public perceptions of AI and socio-technical imaginaries', *AI & Soc*, vol. 38, no. 2, pp. 443–458, Apr. 2023, doi: 10.1007/s00146-022-01422-1.

[93] M. Mazzone and A. Elgammal, 'Art, Creativity, and the Potential of Artificial Intelligence', *Arts*, vol. 8, no. 1, Art. no. 1, Mar. 2019, doi: 10.3390/arts8010026.

[94] S. G. Chiarella, G. Torromino, D. M. Gagliardi, D. Rossi, F. Babiloni, and G. Cartocci, 'Investigating the negative bias towards artificial intelligence: Effects of prior assignment of AI-authorship on the aesthetic appreciation of abstract paintings', *Computers in Human Behavior*, vol. 137, p. 107406, Dec. 2022, doi: 10.1016/j.chb.2022.107406.

© –2023– SESAR 3 JU

EUROPEAN PARTNERSHIP

Co-funded by the European Union

[95] K. Starcke, B. Tuschen-Caffier, H.-J. Markowitsch, and M. Brand, 'Skin conductance responses during decisions in ambiguous and risky situations in obsessive-compulsive disorder', *Cogn Neuropsychiatry*, vol. 14, no. 3, pp. 199–216, May 2009, doi: 10.1080/13546800902996831.

[96] E. Modica *et al.*, 'Neurophysiological Responses to Different Product Experiences', *Computational Intelligence and Neuroscience*, vol. 2018, p. e9616301, Sep. 2018, doi: 10.1155/2018/9616301.

© –2023– SESAR 3 JU

List of acronyms

Table 4: List of acronyms.

Acronym	Description
ATC	Air Traffic Control
RDT	Remote Digital Tower
ATCO	Air Traffic Control Officer
RTO	Remote Towers Operator
MML	Multimodal Machine Learning
CNN	Convolutional Neural Networks
HCAI	Human-Centred Artificial Intelligence
HCXAI	Human-Centred Explainable Artificial Intelligence
NLP	Natural Language Processing
GAN	Generative Adversarial Networks
RNN	Recurrent Neural Networks
XAI	Explainable Artificial Intelligence
LLM	Large Language Model
HAIT	Human-AI Teaming
TIS	Trusted Intelligent System
UAVs	Unmanned Aerial Vehicles
hAIi	Human Artificial Intelligence Interaction
MRTOs	Multiple Remote Tower Operations
pBCI	Brain-Computer Interface
EEG	Electroencephalograms
ECG	Electrocardiograms
SoTA	State-of-the-art
ABSR	Automated Speech-Based Service Requests
NN	Neural Networks
RF	Random Forest
SVM	Support Vector Machine

© –2023– SESAR 3 JU
EUROPEAN PARTNERSHIP
Co-funded by the European Union