Advanced Automation in Aviation

Deliverable ID:	**D2.1**
Project acronym:	**HUCAN**
Grant:	**101114762**
Call:	**HORIZON-SESAR-2022-DES-ER-0**
Topic:	**HORIZON-SESAR-2022-DES-ER-01-WA1-2**
Consortium coordinator:	**Deep Blue**
Edition date:	**28 February 2024**
Edition:	**01.00**
Status:	**Official**
Classification:	**PU**

Abstract

The technological landscape has undergone a significant transformation, with the proliferation of automated systems capable of analysing complex situations, learning, and making intelligent decisions. Driven by advancements in Artificial Intelligence (AI) and Machine Learning (ML), these systems have demonstrated their efficacy in cognitive tasks once thought to be exclusive to humans. In the aviation sector, integrating higher levels of automation and AI into Air Traffic Management (ATM) operations promises improved safety, efficiency, and reliability. This document provides a comprehensive review of advanced automation in the EU Digital Strategy for Mobility, particularly focusing on aviation and ATM. Despite opportunities such as continuous learning and adaptation, challenges like validating and certifying systems, ensuring transparency, and addressing human-centric design concerns persist. Maintaining a human-centric approach, as advocated by the EASA AI Roadmap, is crucial for ensuring that AI enhances human capabilities rather than replaces them entirely.

Authoring & approval

Author(s) of the document

Organisation name	Date
Mohsan Jameel | DLR	19.10.2023
Elisa Spiller | DBL	11.12.2023
Gabriella Gigante | CIRA	14.12.2023
Marco-Michael Temme | DLR	15.12.2023
Lothar Christoffels | DLR	15.12.2023
Jana Meier | DLR	15.12.2023
Henk Hesselink | NLR	16.02.2024

Reviewed by

Organisation name	Date
Elisa Spiller | DBL	13.02.2024
Mariken Everdij | NLR	15.02.2024
Henk Hesselink | NLR	15.02.2024
Sybert Stroeve | NLR	15.02.2024
Jin Choi | NLR	16.02.2024
Marco-Michael Temme | DLR	23.02.2024
Mariken Everdij | NLR	23.02.2024
Federico Galli | EUI	23.02.2024
Giuseppe Contissa | EUI	23.02.2024

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

Organisation name	Date
Paola Lanzi | DBL	29.02.2024
Giuseppe Contissa | EUI	29.02.2024
Mariken Everdij | NLR	29.02.2024
Mohsan Jameel | DLR	29.02.2024
Gabriella Gigante | CIRA	29.02.2024

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

Edoardo Fornaciari | D-Flight	29.02.2024

Rejected by[2]

Organisation name	Date

Document history

Edition	Date	Status	Company Author	Justification
00.01	19.10.2023	Draft	Mohsan Jameel | DLR	TOC
00.02	11.12.2023	Draft	Elisa Spiller | DBL	Drafting (Section 2)
00.03	14.12.2023	Draft	Gabriella Gigante | CIRA	Drafting (Section 3.2)
00.04	15.12.2023	Draft	Mohsan Jameel | DLR Marco-Michael Temme | DLR Jane Meier |DLR	Drafting (Section 3.1)
00.05	26.12.2023	Draft	Gabriella Gigante | CIRA	Drafting (Section 3.2)
00.06	26.12.2023	Draft	Mohsan Jameel | DLR	Drafting and Merging
00.07	16.02.2024	Draft	Mohsan Jameel | DLR Gabriella Gigante | CIRA Marco-Michael Temme | DLR Lothar Christoffels | DLR Jana Meir | DLR	Drafting (Section 4.3.1), (Section 4.2.2), (Section 4.1), (Section 1), (Section 2.1)
00.08	19.02.2024	Draft	Elisa Spiller | DBL	TOC restructured
00.08	19.02.2024	Draft	Marco-Michael Temme | DLR Mohsan Jameel | DLR Henk Hesselink | NLR Elisa Spiller | DBL	Drafting (Section 1), (Section 2), (Section 6), (Section 5)
00.09	20.02.2024	Draft	Mohsan Jameel | DLR Elisa Spiller | DBL	Final for Review

[2] Representatives of the beneficiaries involved in the project

00.10	27.02.2024	Final	Mohsan Jameel | DLR	Final
01.00	28.02.2024	Final	Paola Lanzi | DBL Elisa Spiller | DBL	Editing and quality check

HUCAN

HOLISTIC UNIFIED CERTIFICATION APPROACH FOR NOVEL SYSTEMS BASED ON ADVANCED AUTOMATION

HUCAN

This document is part of a project that has received funding from the SESAR 3 Joint Undertaking under grant agreement No 101114762 under European Union's Horizon Europe research and innovation programme.

Table of contents

Abstract 1

1 Introduction *11*

1.1 Purposes of the document 11

1.2 Intended readership 11

1.3 Structure of the document 11

1.4 Glossary of terms 12

2 Scope and methodology *14*

2.1 Level of Automation 15

2.2 Research Criteria 16

2.3 AI Methods 18

2.3.1 Traditional AI 19

2.3.2 Machine Learning 20

2.3.3 Evolutionary Algorithms 22

3 Advanced Automation and Artificial Intelligence in transport mode *23*

3.1 Introduction 23

3.2 General trends on AA for mobility and transport 23

3.3 Specific trends in different transport modes 24

3.3.1 Air transport 24

3.3.2 Rail transport 25

3.3.3 Road transport and mobility 26

3.3.4 Maritime transport 27

3.4 Opportunities and challenges 27

3.5 Takeaway messages 29

4 Higher Automation and AI in aviation *30*

4.1 Introduction 30

4.2 Airspace Optimization 31

4.2.1 Dynamic Airspace Configuration 33

4.2.2 Human Autonomy Teaming 43

4.2.3 Planning System Development 44

4.3 Assistant to Human 47

4.3.1 Human Assistant 48

4.4 Explainable Artificial Intelligence (XAI) 70

4.4.1 XAI Terminology 71

4.4.2 XAI Principles and Attributes 73

4.4.3 XAI Taxonomy 74

4.4.4 XAI Methods 76

4.4.5 XAI main Toolboxes 79

4.4.6 XAI from Explanation Interface perspective (Chromik et al., 2021) ... 82

4.5 Performance Measures ... 83

4.5.1 Human metrics ... 83
4.5.2 Machine metrics ... 84
4.5.3 Team metrics ... 84
4.5.4 Metrics classification ... 85

4.6 Current Development in Aviation ... 85

4.6.1 Projects in the last five years ... 85
4.6.2 Opportunities and Challenges ... 100
4.6.3 Conclusion ... 102

5 Opportunities and Challenges ... 103

5.1 Workshop ... 103

5.2 Opportunities and Challenges ... 108

5.2.1 Efficient data processing ... 108
5.2.2 Continuous learning and adaptation (self-learning systems) ... 109
5.2.3 Personalised assistants ... 110
5.2.4 Long-term adaptation between humans and AI ... 110
5.2.5 Support in crisis management ... 111
5.2.6 Enhanced problem-solving ... 111
5.2.7 Dynamic Airspace Reconfiguration ... 112
5.2.8 AI and the greening of aviation ... 113
5.2.9 Conclusion ... 114

6 Summary and Conclusion ... 117

6.1 Summary ... 117

6.2 Conclusion ... 118

References ... 119

List of acronyms ... 127

EUROPEAN PARTNERSHIP

Co-funded by the European Union

List of figures

Figure 1 Key characteristics of an AI system (Stefan van Duin, N. B. 2018) 18

Figure 2 Relation between AI, Machine Learning and underlying methods and infrastructure (Stefan van Duin, N. B. 2018) 21

Figure 3 Classification of Literature survey of Airspace Optimization 32

Figure 4 Generation of sector structure depending on task load in (Ingrid et al. 2018) 36

Figure 5 Example for the overlapping sectors (blue areas, the darker, the more sectors overlap) and the SBBs (bordered by red lines) (Ingrid et al. 2018) 36

Figure 6 Overview of the dynamic sectorization approach taken (Ahrenhold, et. al 2023) 37

Figure 7 Surveillance Display used for DAR Manager and ATCO - Active (Solid) and Planned DAR (Dashed) (Teutsch, et. al.. 39

Figure 8 Information exchange services architecture from between ATM and U-space (López, et al. 2023).. 40

Figure 9 Workflow described (Tang, et al. 2021) 43

Figure 10 Methodological overview of the proposed solution (Kim, et. al 2022) 45

Figure 11 The levels of Assistant on the basis of the output (Grochow, 2020). 49

Figure 12 Perspectives on human-machine collaboration (Xiong et al., 2022) 53

Figure 13 Human AI Teaming concept overview (EASA 2023).. 54

Figure 14 Challenges for Human-Machine Collaboration in risky decision-making (Xiong et al., 2022) 57

Figure 15 Fitts list (Fitts, 1951). 59

Figure 16 New Fitt's list (SEI, 2022) 59

Figure 17 Flowchart with the main ideas of the artificial trust-based task allocation method for a team consisting of one human and one robotic agent. (Ali et al., 2022) 61

Figure 18 Definition of Trust in various literature studies 64

Figure 19 Influencing factors of trust 68

Figure 20 The five main perspectives for the need for XAI (Saeed, W. et al., 2023) 70

Figure 21 Synthesised framework for XAI research from a user perspective (Haque, A. B, et al., 2023) 71

Figure 22 XAI Concept (DARPA-BAA-16-53, 2016) 72

Figure 23 Illustration of the four principles of explainable artificial intelligence according to (Phillips, P. et al., 2020)........ 73

Figure 24 XAI taxonomy according to the Functioning based approach proposed by Samek and Muller (Speith, T., 2022) 74

Figure 25 XAI taxonomy according to the Result based approach proposed by McDermid et al (Speith, T., 2022)........ 74

Figure 26 XAI taxonomy according to the mixed based approach (Speith, T., 2022) 75

Figure 27 Overview of a complete taxonomy (Schwalbe, G., et al., 2023) 76

Figure 28 Chronology of the development of successive explanatory methods (Holzinger, A. et al., 2020)........ 76

Figure 29 XAI methods with associated types of explanations (N: Numeric, R: Rules, T: Textual, V: Visual), stage (Ah: Ante-hoc, Ph: Post-hoc), scope (L: Local, G: Global) of explainability, and the design spaces (P: Prediction, O/A: Optimisation/Automation, A: Analysis, M/S: Modelling/Simulation) (Degas, A., et al., 2022) 78

Figure 30 Number of stars on GitHub for the most popular repositories (Holzinger, A. et al., 2020).. 80

Figure 31 Psychological model of explanation. Yellow boxes illustrate the underlying process, the green boxes illustrate the measurement opportunities and the white boxes illustrate potential outcomes (Gunning, D et al., 2021) 80

Figure 32 Measure of Explanation proposed by DARPA (DW, 2019)........ 81

Figure 33 Explanation Interface types (Gunning, D et al., 2021) 82

Figure 34 Explainable AI market forecast........ 101

Figure 35 WP 2 workshop schedule and goals........ 103

Figure 36 Opportunities (1/4) collected and discussed during the HUCAN WP 2 workshop 104

Figure 37 Opportunities (2/4) collected and discussed during the HUCAN WP 2 workshop 105

Figure 38 Opportunities (3/4) collected and discussed during the HUCAN WP 2 workshop 105

Figure 39 Opportunities (4/4) collected and discussed during the HUCAN WP 2 workshop 106

Figure 40 Classification of opportunities into four quadrants according to importance generated during the HUCAN WP 2 workshop 106

Figure 41 Collection of challenges (1/2) associated with opportunities discussed during the HUCAN WP 2 workshop........ 107

Figure 42 Collection of challenges (2/2) associated with opportunities discussed during the HUCAN WP 2 workshop........ 108

List of tables

Table 1 Glossary of terms 12

Table 2 Summary of the differences in AI capabilities and nature of each category 19

Table 3 Overview of the literature survey, with types, subtypes and title 32

Table 4 Airspace sectorization classification criteria provided in (Flener,P. & Pearson, J. 2013) 34

Table 5 Different terms for man machine interaction 50

Table 6 List of factors influencing trust in AI 65

Table 7 Explainable AI Methods list and repositories 77

Table 9 Summary of Opportunities and Challenges 114

Table 10 List of acronyms 127

1 Introduction

1.1 Purposes of the document

This document covers deliverables **D2.1** (Advanced automation in aviation: current and future developments, opportunities and challenges) of the HUCAN project. D2.1 reflects on the output of **Task 2.1** (Advanced automation and artificial intelligence in transport modes), **Task 2.2** (Advanced automation in aviation: current developments and future scenarios) and **Task 2.3** (Opportunities and Challenges identification). This document will serve as one of the inputs to **Task 4.1** (Case studies introduction: level of automation analysis and certification issues).

This document covers the **OBJ1** - Landscape of advanced automation within the EU Digital Strategy for Mobility and Air Traffic Management (ATM) of the HUCAN project, by providing following,

- review and consolidation the levels of automation taxonomy
- a comprehensive scientific review of the state-of-the-art of advanced automation in transport and Aviation,
- discussed opportunities to apply advanced automation and Artificial Intelligence (AI) in ATM
- identify challenges in achieving advanced automation and impact of AI technology

1.2 Intended readership

The intended audience for this document is:

- Single European Sky ATM Research Programme (SESAR) research networks
    - particularly, projects under the SESAR Joint Undertaking (SJU) flag "*capacity-on-demand and dynamic airspace*"
- EU and national representative regulatory authorities and policymakers
- Air Navigation Service Providers (ANSPs) and industrial stakeholders
- academic community

1.3 Structure of the document

The report is structured into five chapters.

The first chapter provides an overview and explains the structure of the document.

Chapter 2 provides a discussion of the level of automation, which is important for categorising automation systems. It discusses various taxonomies available in various fields and, in particular, aviation, such as SESAR and European Union Aviation Safety Agency (EASA). It then provides a taxonomy that is used in this document and a recommendation to streamline it across different entities. After that, this chapter lists the research criteria used to perform the literature survey, which includes choices made in including and excluding certain areas. Lastly, the chapter discusses various AI methods available and explains different categorizations. It further illustrates the range and variety of AI methods.

Co-funded by the European Union

Chapter 3 provides a thorough literature survey of the advancement in different transport modes, including air transport, rail transport, road transport, and maritime transport. The chapter starts with a discussion of general trends in mobility and then discusses specific trends related to each mode of transport in detail.

Chapter 4 focuses on two main topics related to aviation: airspace optimization and human assistance in connection with higher automation. The chapter provides a comprehensive literature survey of the current trends and advancements in these topics, which directly relate to the SESAR flagship "capacity-on-demand and dynamic airspace" and the use cases defined in the HUCAN project. Some of the highlights of this chapter include a discussion about technical systems, eXplainable AI and past projects in the areas.

Chapter 5 discusses future opportunities for high aviation automation and its associated challenges. It presents the results of the workshop on "Opportunities and Challenges" organised as part of Work Package 2.

Finally, **Chapter 6** concludes the document with a summary of the main points covered in the preceding chapters.

1.4 Glossary of terms

Table 1 Glossary of terms

Term	Definition	Source of the definition
Advance Automation	It refers to the use of a system that, under certain conditions, operates without direct human intervention.	ISO/IEC 22989:2022(en), 3.1.7
Air Traffic	All aircraft in flight or operating on the manoeuvring area of an aerodrome.	ICAO Annex11 - ATS
Artificial Intelligence	“The branch of computer science that deals with the development of computer systems capable of performing tasks that typically require human intelligence. These tasks include learning, reasoning, problem-solving, perception, natural language understanding, and interaction with the environment." “Technology that can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with”.	[Russell 2010] EASA AI Roadmap 2.0
Air Traffic Management	The dynamic, integrated management of air traffic and airspace including air traffic services, airspace management and air traffic flow management - safely,	ICAO 4444 - ATM

	economically and sufficiently - through the provision of facilities and seamless services in collaboration with all parties and involving airborne and ground-based functions.	
High Automation	Automation supports the human operator in information acquisition and exchange, information analysis, action section and action implementation for all tasks/functions. Automation can initiate actions for most tasks. Adaptable/adaptive automation concepts support optimal socio-technical system performance.	SRIA 2020
Explainable AI (XAI)	Explainable AI refers to the capability of AI systems to provide understandable explanations of their decisions and actions to human users. It aims to enhance transparency, trust, and accountability in AI systems by making their internal mechanisms and reasoning processes interpretable and comprehensible to users. XAI techniques enable users to understand how AI systems arrive at their outputs, which is crucial for building trust, verifying correctness, detecting biases, and identifying potential errors or limitations in AI-driven decisions.	Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)
Digital Assistant	A concept that includes Artificial Intelligence, and goes beyond tools based on machine learning algorithms that provide data and information to a human operator.	

2 Scope and methodology

Over the past decade, the technology landscape has undergone a remarkable transformation, paving the way for the expanded role of automation. One of the most significant developments has been the evolution of these systems' capabilities to analyse complex situations, learn from them, and make intelligent decisions autonomously. This progression has exceeded expectations, as high automation systems have demonstrated proficiency in cognitive tasks previously believed to be exclusive to human capabilities. The transformation can largely be attributed to recent advancements in Artificial Intelligence and Machine Learning technologies, evidenced by applications like ChatGPT, AlphaGo, and AlphaFold, to name a few.

The field of ATM has seen rapid changes and growth, with increased demand and the emergence of new players such as Unmanned Aerial Vehicles (UAVs). To address these challenges and meet future needs, it is expected that the integration of higher levels of automation in ATM systems will be critical. Strategic Research and Innovation Agenda (SRIA) defines high automation as

> “Automation supports the human operator in information acquisition and exchange, information analysis, action section and action implementation for all tasks/functions. Automation can initiate actions for most tasks. Adaptable/adaptive automation concepts support optimal socio-technical system performance.” (SRIA 2020)

By leveraging highly automated systems, the ATM industry can potentially transform how it operates with improvements in safety, efficiency, and reliability. Automated systems can analyse large amounts of data from various sources, including radar, weather sensors, and flight plans, to provide controllers with real-time insights and decision support. With the use of machine learning algorithms, these systems can adapt to changing conditions and optimise air traffic flow dynamically, leading to smoother operations and fewer delays. Additionally, this technology can assist air traffic controllers and pilots by reducing their workload and alleviating stress associated with their responsibilities. Furthermore, increasing the levels of automation in ATM systems can augment human capabilities rather than replace them entirely, which is clearly expressed in EASA AI Roadmap (EASA, 2023) as a human-centric approach.

Despite the clear benefits and progress made in automation technology, there remain significant challenges surrounding the integration of higher levels of automation and artificial intelligence in ATM operations. The SIRA for Digital European Sky, issued by SJU, highlights numerous critical issues and obstacles. SIRA emphasises the need to concentrate on developing new methodologies for validating and certifying advanced automation that ensure transparency, legal compliance, robustness, and stability under all conditions while taking into account a future ATM environment that relies on multiple AI-based systems of systems, with a focus on human-centred design.

Aligned with the project's objectives, this research endeavours to facilitate the utilisation of novel systems that offer discernible benefits in terms of effectiveness and efficiency. The primary focus of this document is to investigate the opportunities and challenges associated with advanced automation in the field of transportation, particularly within aviation. These challenges encompass concerns related to human-machine interactions and the organisational impact of automation (Lim (2023) and Fortunati & Edwards (2022)). Additionally, paramount to the success of these advanced systems is the assurance of security and resilience against cyber threats (Lee (2023)). Striking a delicate balance between automation and human involvement, especially in decision-making processes, poses a multifaceted challenge (Samad (2023)). The objective of this document is therefore to provide a thorough overview of the concrete needs that exist in the transport sector, in particular in the aviation sector, for the implementation of advanced automation. The aim is to gain a contextual understanding of the requirements for standardising regulatory frameworks and adapting certification methods to ensure the safe and responsible use of advanced automation technologies.

2.1 Level of Automation

"Automation" is generally described as the use of control systems and information technology to minimise the need for human input, particularly in repetitive tasks. Accordingly, "Advanced Automation" (AA) (ISO/IEC 22989, 2022) refers to the use of a system that, under certain conditions, operates without direct human intervention. These definitions highlight the evolution and sophistication of technological systems in reducing human involvement in certain processes, from basic automation for repetitive tasks to more advanced automated systems capable of operating independently.

The SESAR research and innovation initiative has been instrumental in supporting air traffic controllers and reducing their workload to enhance the efficiency of the Air Traffic Management (ATM) system. While there is a shared understanding that the future of ATM will involve higher levels of automation, a collective vision is imperative to shape a research roadmap detailing specific actions. In line with this, the most recent directives (SESAR, 2020; SESAR, 2023(a); SESAR, 2023(b)) emphasise a comprehensive examination of automation characteristics and the establishment of conditions to facilitate its practical and scalable implementation, with a keen consideration of certification aspects.

As evidenced in the literature over time (Sheridan et al., 1978; Parasuraman et al., 2000; Dekker et al.,2002; Save et al., 2012) advances in technology have provided increasingly sophisticated ways to automate human operator tasks, thereby enhancing human-machine performance within complex systems. In this regard, the concept of automation is seen as nuanced, rejecting the binary notion of 'all or nothing'. Instead, it emphasises the importance of deciding the degree to which a task should be automated. Beyond the mere delegation of tasks to machines, the introduction of automation implies qualitative changes in human practices. Recognising this, our approach involves considering different levels of automation within each function to establish guidelines for effective automation solutions.

Accordingly, in the context of the SESAR level of automation taxonomy (LOAT) approach (SESAR, 2013), the key is to determine the degree to which automation should be implemented, recognising a wide range of options between these extremes and carefully evaluating the associated advantages and disadvantages. Qualitatively, high-level automation support for information acquisition involves integrating data from different sources, filtering and highlighting relevant information based on predefined criteria visible to the user. Similarly, high-level automated support for information analysis assists users in comparing, combining and analysing information items, and triggers alerts when attention is required. High-level automated decision-making means that the system autonomously generates options and decides on actions, with human notification only on request. Support for the execution of action sequences, both automatic and user-initiated, is an integral part of high-level automation, enabling monitoring and intervention as required.

More recently, the EASA AI roadmap (EASA, May 2023) delves into crucial aspects of artificial intelligence (AI) and autonomy within the aviation sector. According to the LOAT approach, the roadmap emphasises the importance of adaptivity in the learning process, enabling performance improvement through experience, particularly in machine learning contexts like online learning.

In particular, the new roadmap provides a new perspective on categorising the level of automation associated with AI-based technology. There are three scenarios for classifying human interaction with machines: human assistance (Level 1), human-AI teaming (Level 2), and advanced automation (Level 3 AI). Specifically, Level 1 AI supports human augmentation (L1A) or human cognitive assistance in decision-making and action selection. Level 2 AI is further subdivided into cooperation (Level 2A) and collaboration (Level 2B), characterised by the type of interaction and shared awareness between humans and AI-based systems. Level 3 AI introduces distinctions between 3A and 3B, where 3A involves supervised automatic decision-making and action implementation, while 3B involves unsupervised automatic decision-making and action to support safety, especially in the absence of human supervision (EASA, February 2023; EASA, May 2023).

For the purposes of HUCAN, it is worth to be noted that the classification also introduces the distribution of authority (EASA, February 2023), ranging from full authority for the end-user (up to Level 2A AI), through partial authority (Level 2B AI), to full authority for the AI-based system (Level 3 AI). As a result, the Human-AI Teaming (HAT) further classifies the intensity of the interaction, distinguishing between cooperation (Level 2A) with a directive approach and collaboration (Level 2B) with a focus on joint problem-solving and shared awareness. This nuanced approach provides insight into the collaborative dynamics between humans and AI-based systems in aviation operations, taking into account different levels of authority and communication requirements.

In light of the above, for the purposes of HUCAN, advanced automation is intended as the combined utilisation of sophisticated technologies, often incorporating AI, machine learning (ML), and robotics, to enhance and streamline complex processes in various industries (Baribieri et al., 2022). In this realm, automation goes beyond basic, rule-based tasks, supporting the human operator's cognitive capacities in information acquisition and exchange, information analysis, action selection and action implementation, also exhibiting a higher level of adaptability and autonomy (SESAR, 2020). In this regard, AA systems can analyse large datasets, learn from experiences, and make intelligent decisions, allowing them to operate in dynamic and unpredictable environments. These systems often involve interconnected components, such as sensors, actuators, and computing systems, working together to optimise efficiency, reduce human intervention, and achieve higher levels of precision and reliability.

2.2 Research Criteria

Within the framework of the HUCAN project, diverse research criteria have been strategically employed to guide investigations and analyses.

The primary aim of this document is to conduct a thorough analysis of the present and future applications of advanced automation and AI within various transportation sectors, encompassing automotive, trains, ships, and aviation. This analysis seeks to uncover how automated systems and AI technologies are currently utilised and planned for implementation across different transport domains. The emphasis is on identifying both commonalities and distinctions in the adopted solutions while comprehending the operational-level benefits and challenges. In this regard, in the first part of the report, the focus is primarily on the European context, aligning with the policy objectives and funding programs advocated by the European Union in the transport sector. Particular attention is devoted to developments in the aviation sector, recognising its significance as a focal point for advancements in high-level automation. The methodology employed for scoping the literature review is structured to leverage insights from previous research initiatives funded under the H2020 and HorizonEU programs. The scope encompasses artificial intelligence, advanced automation, mobility, and transport, with a territorial focus on the EU and the EEA.

Subsequently, the attention turns specifically to the aviation sector, building upon the findings related to the challenges and opportunities of advanced automation in diverse transportation modes. The study aims to delineate the current state of developments and future scenarios of advanced automation and AI in the aviation sector. Considerations include SESAR expectations, outcomes from prior SESAR exploratory research, and initiatives explored by industry and ANSPs for ATM-related air and ground systems in the short, medium, and long term. The research criteria encompass "high level of automation in aviation" and "artificial intelligence in aviation," refining the focus on aviation-specific contexts, exploring applications of automation and artificial intelligence, and delving into the ATM control phase, with a specific focus on the working environment and tasks of air traffic controllers. The inclusion of a "project of interest" criterion allows for targeted analysis of specific projects with potential impacts on the broader research area.

Furthermore, a systematic literature review has been conducted. The review process may have different biases that can affect the effectiveness of the research:

- ✓ Reading before planning (defining a review protocol that specifies the research question being addressed)
- ✓ Reading everything / read unlinked papers (detect as much of the relevant literature as possible)
- ✓ Reading outdated version of a paper/book
- ✓ Reading but not writing
- ✓ Start reading with few resources
- ✓ Language bias
- ✓ Not keeping bibliographical information

In order to mitigate the above-cited biases, the review has been performed according to a defined process. The process has been set up integrating the specific needs within a typical framework of a “systematic literature review”.

The process includes the following phases:

a) Planning the review
b) Conducting the review
c) Writing the review

Such phases are executed in cascade and each phase is organised according to a series of steps.

Planning the review is the first phase agreed among the partners. It has preliminary identified the need of the review and then has defined a precise review protocol. The review protocol mainly relies on three key aspects: (i) formulate appropriate research questions; (ii) identify the most appropriate temporal frame; (iii) identify the sources to consider.

Co-funded by the European Union

The temporal frame has been set equal to the last five years (2018 – 2023). The sources point out where to find literature. They have been identified categorising the main purposes of the studies on the problem of interest. Apart from the projects, the main considered sources have been Google Scholar, Research Gate, Science Direct, AIAA ARC, Journals such as International Journal of Information Technology & Decision Making.

Three major criteria for selection have been fixed:

- Relevance
    - To what extent the material covers the research questions?
    - Does it provide sufficient details to gain a clear picture of the results achieved?
    - Does it overlap with other research?
- Authority
    - Has it been published by a reputable source or is it possible to justify why it is an important source?
- Temporal horizon
    - Is the material still influential in the field?
    - Is it keeping up to date with new research?

2.3 AI Methods

Advancement in the field of AI is the dominating factor in revolutionising the technology landscape. From autonomous driving to language translation and social networking, AI has made higher automation possible in almost all domains. The power of AI models and techniques has opened up endless possibilities, making it the go-to technology for implementing automation and autonomous systems. Therefore, it is important to have an overview of AI technologies and methods. This section covers the various methods and techniques available under the AI umbrella. Our aim is to provide an overview, with additional references available for those who wish to learn more.

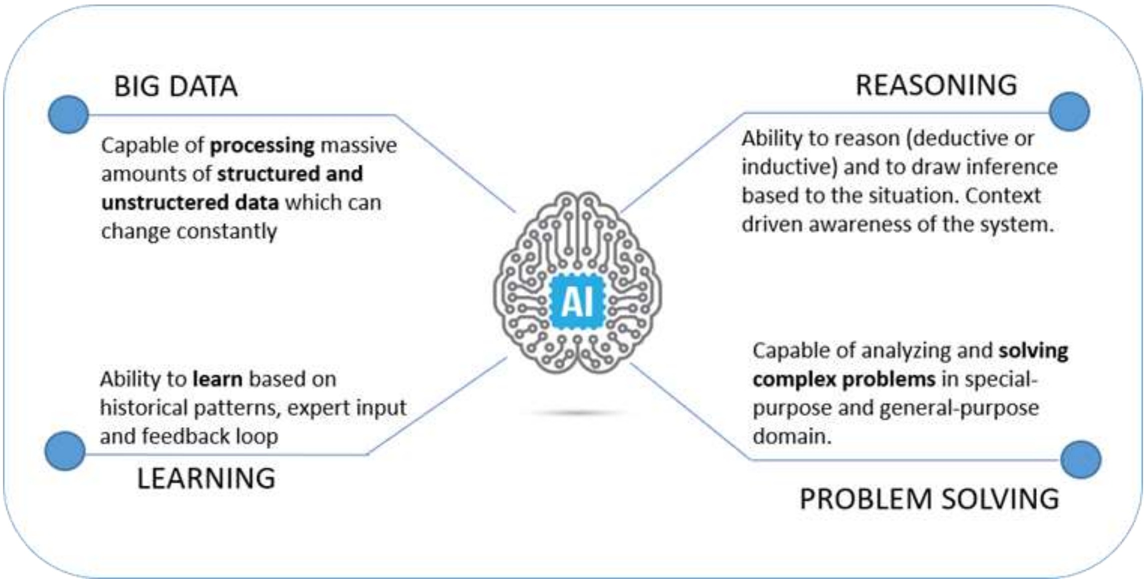


Figure 1 Key characteristics of an AI system (Stefan van Duin, N. B. 2018)

AI is a multidisciplinary field that seeks to develop intelligent systems capable of performing tasks that typically require human intelligence. These tasks encompass a wide range of domains, including problem-solving, decision-making, perception, language understanding, and learning. AI systems use computational methods and algorithms to mimic or replicate human cognitive functions such as reasoning, learning, planning, and perception. From a technical standpoint, AI can be defined as,

> "[T]he branch of computer science that deals with the development of computer systems capable of performing tasks that typically require human intelligence. These tasks include learning, reasoning, problem-solving, perception, natural language understanding, and interaction with the environment." (Russell, 2010)

From an operative standpoint, EASA has defined AI in its AI Roadmap 2.0 as

> "Technology that can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with." (EASA 2022)

There are various methods that constitute AI and can be grouped into three broader categories: traditional AI, machine learning and evolutionary algorithms. Table 1 provides a short summary of the differences in capabilities and nature of each category.

Table 2 Summary of the differences in AI capabilities and nature of each category

Characteristic	Traditional AI	Machine Learning	Evolutionary Algorithms
Deterministic	Yes	No	No
Knowledge Engineering	Extensive	Not Required	Not Required
Data-Driven	No	Yes	No
High-Dimensional Search	No	Yes	Yes
Adaptable	No	Yes	Yes
Interpretability	High	Varies	Varies
Scalable	No	Yes	Yes
Generalisation	No	Yes	No

2.3.1 Traditional AI

Traditional AI, also known as classical AI, is the term referred to pre-modern machine learning techniques. The techniques that fall under this category involve explicitly programming rules and logic to imitate human intelligence. They were popular methods due to their deterministic nature, high interpretability and ease of implementation. Rule-based systems, Symbolic AI, Knowledge Engineering and Expert Systems are a few examples in this category. However, these methods exhibit limited scalability, adaptability and generalisation capabilities, limiting their ability to model and capture complex processes. Some examples of traditional AI include rule-based systems, symbolic AI and Knowledge Representation.

- **Rule-Based Systems**

  Rule-based systems operate on predetermined rules and logic, meaning the output is solely determined by the input and rules. These systems are deterministic, as they produce the same output for the same input each time. Rule-based systems have seen extensive use in air traffic control for various applications, including managing aircraft flow by defining strict rules and procedures for routing, landing/takeoff sequencing, and aircraft separation (Buchanan, et al 1984). An example of rule-based system by CANSO "Rule-Based Systems in Air Traffic Control" (CANSO 2019)

- **Symbolic AI**

  Symbolic AI employs symbols and logic to represent knowledge and perform reasoning, following deterministic rules for logical deduction and inference. This technique is used in flight planning systems, where logical rules and representations of flight constraints are used to determine optimal routes, fuel consumption, and flight schedules. "Symbolic AI for Flight Planning " (Bazzan, et al. 2014).

- **Knowledge Representation**

  Knowledge representation involves structuring knowledge in a format easily processed by computers. Brachman and Levesque's (Brachman, et al. 2004) work provides an insight into this area, explaining techniques and frameworks for effective knowledge representation and reasoning.

2.3.2 Machine Learning

The recent growth and progress in AI can be largely attributed to the advancements in machine learning techniques and methods. Machine learning is a field that focuses on computer programs' ability to learn patterns and relationships from past data, and then use that information to make decisions on new and unseen data. Unlike traditional AI, machine learning uses data-driven approaches to learn patterns and make predictions without explicit programming of rules. This allows for better generalisation, adaptation to new scenarios, and scaling up to larger problems. These methods excel at finding solutions in high-dimensional search spaces, allowing them to model very complex problems and tasks with ease. However, due to the high dimensionality of the search space, interpreting the decision can be challenging. Additionally, these methods can exhibit non-deterministic behaviour due to factors such as random initialization or stochastic optimization techniques. These techniques include a wide range of algorithms and methodologies, such as neural networks, decision trees, support vector machines, and more.

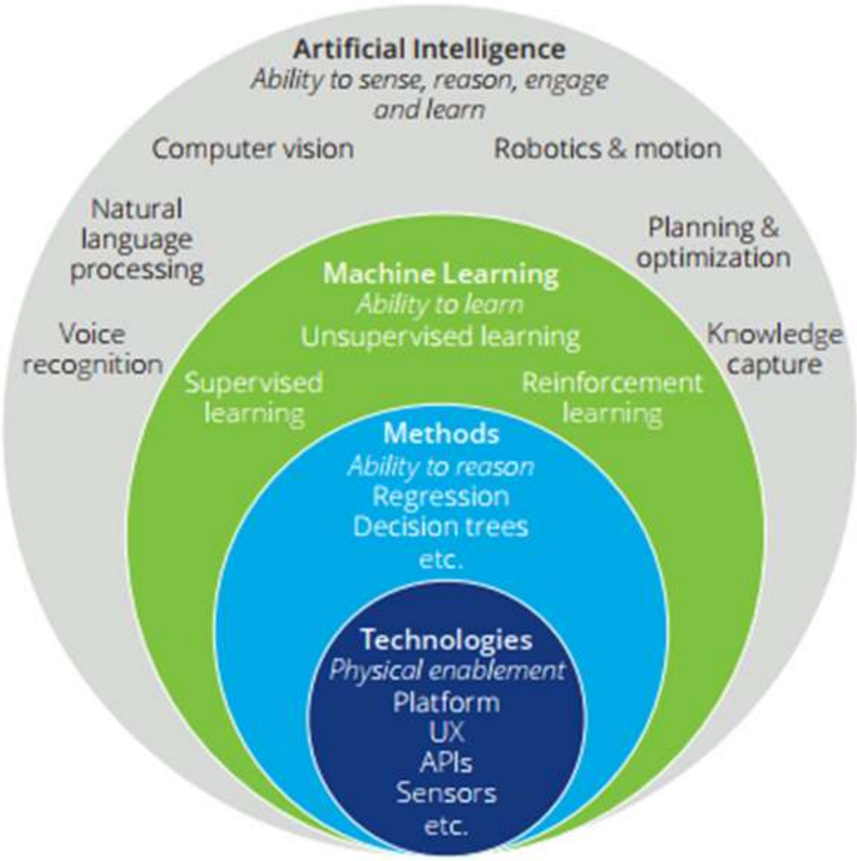


Figure 2 Relation between AI, Machine Learning and underlying methods and infrastructure (Stefan van Duin, N. B. 2018)

Machine Learning is further divided into three main learning paradigms, i.e. supervised learning, unsupervised learning, and reinforcement learning. In Supervised Learning paradigms, each instance of data consists of observations and explicit labels. This approach is similar to learning from examples. The learning algorithm takes guidance from the labelled examples and corrects its decision logic. It is the most effective way of learning since the target is well-defined. Some examples of supervised learning include Deep Neural Networks, Regression, Decision Trees, Transformers and more. However, one of the major challenges in supervised learning is generating labelled data, which is expensive to collect and requires additional effort.

On the other hand, unsupervised learning paradigms define techniques to learn from unlabeled data, for example, clustering similar data into coherent groups. It allows the power to tap into vast amounts of unlabeled data. Examples include K-Means Clustering, Principal Component Analysis, Hierarchical Clustering and more. However, the major drawback in this paradigm is the lack of a well-defined target, which makes it harder to achieve useful results. Therefore, mostly unsupervised learning techniques are used in combination with supervised learning methods, such as feature engineering and additional input signals. This combination of unsupervised and supervised learning techniques has been found to be quite effective in various applications, including image and speech recognition, natural language processing, and recommendation systems.

Finally, reinforcement learning is a type of machine learning approach that is based on learning by experience. In this learning method, an agent interacts with an environment by taking certain actions and receiving feedback from the environment in the form of the next state of the environment and a reward. The reward that the agent receives provides feedback about the impact of the action taken by the agent, whether it improves the environment or makes it worse. There are several examples of reinforcement learning algorithms such as Deep Q-Networks (DQN), Proximal Policy Optimization (PPO), and Asynchronous Advantage Actor-Critic (A3C), among others. These algorithms are designed to enable the agent to learn from experience to make better decisions in the future. Reward shaping is considered a critical process in the learning process as it represents a signal to the agent about what it should learn. However, an incomplete reward can lead to unintended behaviour by the agent. Therefore, it is essential to design proper reward-shaping techniques to ensure that the agent learns the desired behaviour and avoids unintended behaviour. For a thorough understanding, we recommend following references. (Murphy, K. P. 2012) (LeCun, et al. 2015) (Goodfellow, et al. 2016) (Sutton, et al. 2018).

2.3.3 Evolutionary Algorithms

Evolutionary algorithms are a subfield of computational optimization techniques that fall under the larger umbrella of AI. These algorithms take inspiration from the principles of natural selection and are considered a type of probabilistic optimization tool. Evolutionary algorithms, such as genetic algorithms, evolutionary strategies, and genetic programming, are known for their non-deterministic nature and their ability to use randomization and selection mechanisms that can lead to varying outcomes for the same input or initial conditions. At their core, these algorithms function by generating a population of candidate solutions (often represented as individuals or chromosomes), and then applying mechanisms such as selection, crossover (recombination), and mutation to evolve and refine these solutions over multiple generations. A fitness function determines the quality of the intermediate solutions in the selection process and the final solution. Through this iterative process, evolutionary algorithms aim to discover optimal or nearly optimal solutions to complex optimization problems. They are adaptable, scalable and able to model complex processes as they also find solutions in high-dimensional search space. However, they provide limited ease of solution interpretability and do not generalise to other problems. The fitness landscape provided by data does not guide the evolutionary algorithms' fitness function. It can be considered a strength, allowing it to expand learning exploration and, on the other hand, a weakness as it requires much more computational time to reach the optimal solution.

Overall, evolutionary algorithms are a powerful tool for solving complex optimization problems in a wide range of fields, including engineering, finance, and bioinformatics. They have proven to be a highly effective approach for generating high-quality solutions to difficult optimization problems that would be impractical or impossible to solve using traditional optimization techniques. (Dasgupta et al., 2013) (Bäck et al, 1993).

3 Advanced Automation and Artificial Intelligence in transport mode

3.1 Introduction

The challenges emerging in managing Advanced Automation (AA) greatly depend on the specific characteristics of individual sectors. The HUCAN project addresses the social and technical issues currently experienced in the certification of advanced automated solutions in aviation and aims at developing a holistic methodology and an operational design toolkit to tackle the emerging challenges in this realm. However, this technological transformation is driving a transformative phase for the whole transport and mobility sector. This section provides an overview of the current research and innovation trends in the transport and mobility sector to identify similarities and differences and explores the opportunities and challenges associated with the development and implementation of these solutions at the operational level.

3.2 General trends on AA for mobility and transport

Advanced Automation is a key force in the transformative evolution of industries, improving productivity, quality and safety while ensuring competitiveness in the technology-driven landscape. This technical section examines the research initiatives driving innovation in intelligent mobility and autonomous transport systems, focusing on the integration of AA systems, artificial intelligence and robotics.

Collaboration among academia, industry and government has been instrumental in advancing automation. Research focuses on the development of autonomous systems for management, logistics and transportation, with a strong emphasis on improving efficiency, sustainability and flexibility. The growing emphasis on human-robot collaboration underscores the importance of advanced automation systems working synergistically with human operators to improve productivity and safety.

Considering the scope and the objectives of HUCAN, it is important to stress how the EC recognises AA as a crucial driver of the digital transition, not only as a stand-alone challenge but also as an indispensable catalyst for achieving specific and overarching sectoral goals (EC Competence Centre on Foresight, 2022). By anticipating future mobility trends, the EU aims to create favourable conditions for the development and validation of new technologies and services (EC, 2020). More specifically, AA in transport and mobility will foster critical sustainability goals, optimising the efficiency of transport networks and making a significant contribution to the transition to greener and more environmentally friendly transport solutions and infrastructure (EC, 2023) (EP, 2021a) (EP, 2021b). In this regard, the EC policy strategy emphasises the importance of enabling testing and experimentation and of making the regulatory environment fit for innovation to support the deployment of solutions on the market (EC, 2020).

From a research and innovation perspective, academia and industrial stakeholders are intensifying efforts in common application trends. In autonomous vehicles, significant strides have been made by industry leaders like Tesla and Waymo, showcasing self-driving cars equipped with advanced sensors, cameras, and AI algorithms (Iclodean et al., 2023). Similarly, the freight industry is experiencing a transformation with the introduction of autonomous trucks, promising efficiency gains and reduced labour costs (Dekhtyaruk, 2023).

Automation facilitates predictive maintenance, monitoring, and forecasting the maintenance needs of vehicles and infrastructure (Giordano, et al., 2022). This approach minimises downtime and enhances operational efficiency across various transportation modes.

Traffic management is witnessing a revolution with advanced automation optimising traffic flow and reducing congestion through adaptive traffic signal control systems (Gokasar et al., 2023). The integration of vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication fosters a connected transportation ecosystem, enhancing safety and efficiency on roads (Ahmed et al, 2023).

Automated ride-sharing services are on the horizon, with companies testing prototypes of autonomous taxis. Simultaneously, micro-mobility solutions, such as electric scooters and bikes with automated features, are gaining traction in urban areas, catering to the demand for sustainable and flexible mobility options (Brodersen et al., 2023).

Eventually, in the realm of air transport, urban air mobility (UAM) is emerging with developments in electric vertical take-off and landing (eVTOL) aircraft poised to revolutionise urban transportation. Companies are exploring automated drone delivery systems, presenting innovative solutions for transporting goods in diverse urban and remote areas.

3.3 Specific trends in different transport modes

While macro trends in AA have a universal impact on the mobility and transport sector, their specific manifestations vary between different transport modes. This section presents the different research and development (R&D) pathways within each domain, generally mapping the results obtained in the main projects funded by the EU under Horizon 2020 and Horizon Europe.

3.3.1 Air transport

The aviation industry has seen many technological revolutions, with the smart use of task automation to improve the safety, efficiency and accessibility of air travel. The deep integration of AA and AI stands out as a pivotal force for the safer enhancement of avionics and efficient management of various facets of aviation (EC JRC, 2023).

In this context, AA and AI-based applications are expected to positively **enhance the safety of aircraft, including UAVs and drones**. In particular, these technologies foster safer prototyping and testing of aircraft systems, also contributing to the development of new certification standards (AEROGLASS). In the area of urban air mobility (UAM) these solutions further contribute to improving route planning techniques (LABYRINTH, SAFEDRONE, MONIFLY, COMP4DRONES, TINDAIR and AURORA) and more precise positions of drones and UAS in the U-space (GAUSS), improving sensor performance and connectivity protocols (SAFEDRONE, SAFIR-MED and ASSURED-UAM).

Significant advances are also expected **in ATM, particularly through AI and digital tools that can assist and support ATC decision-making**. These applications can yield major benefits by supporting and accelerating decision-making, reducing workload and enabling controllers to focus on critical tasks by automating less critical and procedural tasks (MAHALO, PJ16 CWP HMI, FARO and SAFECLOUDS.EU).

Researchers are investigating predictive modelling to anticipate safety hazards and collect operational data from flights. Researchers are investigating predictive modelling to anticipate safety hazards and collect operational data from flights. In this regard, it clearly highlights the importance of storing and processing large amounts of operational data to identify patterns and train predictive algorithms for increased safety and efficiency in air traffic management (SAFECLOUDS.EU). Particular attention is also given to the design of the human-machine interface (HMI) to facilitate seamless collaboration in AI-assisted decision-making (PJ16 CWP, HAIKU).

In-flight safety research focuses on the use of advanced sensor technologies and software processing techniques to improve overall safety, also envisioning recommended future requirements (PJ11 CAPITO and ODESSA). More specifically, projects are obtaining intriguing results in improving obstacle detection, avoidance and navigation, especially in challenging conditions such as low visibility or adverse weather (SENSORIANCE, WINFC and VISION). There is also a coordinated effort to monitor the cognitive state of pilots and to assess the impact of highly automated systems on controller performance (STRESS and REPS).

In the area of emergencies, significant research and development efforts are directed towards supporting pilot decision-making in emergencies (SAFENCY project) **and managing onboard pilot incapacitation scenarios** (SAFELAND project). Research efforts are also oriented to improving rescue capabilities in general aviation emergencies, contributing to a holistic approach to safety in aviation emergency scenarios (GRIMASSE).

Finally, **certification projects play a key role in driving innovation in aviation safety**. In this phase, one of the main goals is to improve the assessment of safety areas for both commercial aviation and rotorcraft operations (OPTICS, OPTICS2 and NITROS). They use bottom-up and top-down approaches to assess research maturity and potential real-world applications. These projects are refining the knowledge management framework, using open databases, curating knowledge for innovative training and proposing improved certification processes (ASCOS project).

These aspects will be further addressed in the second section of this document, with a specific focus on AA for airspace optimization and assistance to human operators.

3.3.2 Rail transport

The ongoing digital transformation of rail transport offers a unique opportunity to improve safety and efficiency. AA, as well as increased computing power, artificial intelligence and high-speed wireless connectivity, are driving the adoption of automation in traffic and safety management. This includes precise real-time positioning for concepts such as automatic train operation, virtual coupling and train platooning (EC JRC, 2023).

The EU is emphasising **a unified approach to railway automation based on the European Rail Traffic Management System (ERTMS) system to ensure interoperability**. The transition from outdated GSM technology to 5G-based solutions for safety-critical communications requires careful management to ensure a seamless transition.

Digitalisation is expected to improve rail safety, but also increases **the reliability of infrastructure and rolling stock through continuous monitoring and preventive maintenance**. Innovative sensor systems

Co-funded by the European Union

strategically placed along the tracks enable real-time health monitoring, early fault detection and predictive maintenance. These systems also enhance security by detecting unauthorised intrusions.

Intelligent infrastructure research focuses on future-proof components and improved track systems. Intelligent mobility management initiatives aim to advance automated transport systems within a standardised ICT environment.

Accurate train positioning solutions using advanced GNSS approaches have been developed to cope with growing rail traffic. In particular, the use of AA and AI-based solutions is making a significant contribution to optimising yard operations (OPTIYARD) and real-time planning solutions to minimise delays (ARRIVAL and ON-TIME). Real-time monitoring also optimises network capacity, reduces delays and manages disruptions caused by extreme weather conditions (IN2RAIL).

Finally, data-driven solutions for energy and asset monitoring across the rail system are contributing to a significant leap forward in improving the safety, efficiency and technological performance of rail transport systems (IN2DREAMS).

3.3.3 Road transport and mobility

In road transport, the main research trends focus on human-machine interactions, especially forwarding connected and automated vehicles. From the technical standpoint, particular attention is devoted to the use of automation safety strategies and testing. Efforts also converge on communication standards (EC JRC, 2023).

Recent research activities on **Advanced Driver Assistance Systems (ADAS)** focus on improving driver-vehicle interaction by improving control transfer, testing new sensors and developing algorithms to increase system efficiency and reliability. Challenges include the adaptation of ADAS systems to different driving conditions and the seamless transfer of control between driver and vehicle. Efforts are also focused on assessing driver fitness, fatigue and reaction times to counter risky behaviour and reduce the risk of human error (MEDIATOR, ADASANDME, I-DREAMS and FITDRIVE). Testing also addresses the seamless transfer of control between driver and vehicle, taking into account the driver's state, environmental conditions and accident-prone situations (MEBESAFE).

Other research has worked on **efficient communication among automated systems, drivers and the surrounding environment**, with significant progress in communication standards and algorithms (ENSEMBLE, COSAFE). Sensor systems monitor driver behaviour and enable communication with the environment (HADRIAN and SAFER-LC), infrastructure (Vehicle to Infrastructure; V2I), other vehicles (Vehicle to Vehicle; V2V) and the vehicle environment (SMARTCARS and SAFE STRIP), in particular for the detection of hazardous situations (VI-DAS and DENSE). Some initiatives also focus on improving safety through enhanced interactions between automated vehicles and other road users and facilitating the integration of automated vehicles (TRANSAID), even providing high-impact demonstrations of autonomous minibuses (AVENUE).

Safety testing tools are increasingly based on virtual environments, incorporating various features in digital models of the human body and analysing traffic accident data to develop effective accident prevention strategies (SENIOR, SIMUSAFE). Notable projects improve the safety of vulnerable road users by enhancing dedicated in-vehicle active safety systems and contribute to safer urban planning (PROSPECT, SAFE-UP, XCYCLE, HANDSHAKE).

Finally, some research projects focus on **vehicle automation as well as NTM systems and digital infrastructure for Coordinated, Connected and Automated Mobility (CCAM)**, addressing network prioritisation and traffic orchestration strategies as well as user needs and requirements and human factors issues related to road transport automation and integrated mobility solutions (CONDUCTOR, SINFONICA, ORCHESTRA, FAME).

3.3.4 Maritime transport

Significant progress has been made in the field of maritime safety, through technological innovation and policy improvements (EC, 2020) (EC JRC, 2023).

A key step towards smarter and safer maritime transport is the implementation of the EU VTMIS. This interoperable system will enhance maritime traffic and transport, improving safety, efficiency and response to incidents. Technologies such as external and hull inspections by drones complement this system, streamlining the inspection process and emphasising operational issues over documentation.

Several projects have made a significant contribution to maritime safety, with significant technological **improvements for streamlined ship inspections** (SAFEPEC FP7), **early warning systems in maritime radar surveillance** (RANGER) **and a collision avoidance solution using advanced sensors** (SAFENAV) by developing a prototype using historical and real-time data. Other research initiatives developed e-navigation solutions to improve information sharing in the maritime sector (EFFICIENTSEA 2).

On the other hand, researchers also explored **AA and AI-based solutions to improve evacuation procedures**. Research to enhance smart life jackets, also incorporating wristbands and augmented reality applications aims to redefine evacuation procedures for passenger ships for enhanced situational awareness (SAFEPASS). Significant progress has been made in technologies for tracking passengers and crew during emergency evacuations: localisable life jackets, wristbands with integrated functionalities for specific passenger groups, including localisation radars for people on board lifeboats, people counting handheld devices and intelligent decision support systems (LYNCEUS2MARKET). AI and AR solutions in a massive evacuation vehicle (MEV) are being tested to better support evacuation procedures.

The projects analysed are in line with policy objectives and focus on vessel traffic monitoring, accident investigation and safety data management. To reduce the number of accidents at sea, continued research and innovation are essential, in particular for the integration of sensing, tracking and routing solutions into ships and monitoring systems. **Certification procedures, crew training and regulatory requirements for innovative equipment require further attention**.

3.4 Opportunities and challenges

The research on AA and AI-based solutions promises to significantly improve the safety and resilience of transport and mobility, reducing the environmental impact of travel and better meeting the users' needs and societal expectations.

The analysis of current trends in AA and AI-based solutions in transport and mobility reveals four general drivers: safety, resilience, sustainability and acceptability.

- Safety is a ubiquitous requirement in research and innovation activity on transport and mobility, generally addressing all immediate challenges concerning the safe development and implementation of automation and digitalisation for transport.
- Resilience is a complementary requirement for society to function, as human activities, including commuting and recreation, and supply chains depend on transport. In this context, investments in AA and AI tend to focus on strengthening the resilience of EU transport in times of crisis and improving the cyber-robustness of digital systems.
- Sustainability is a key focus, leveraging AI-driven route planning to create greener transport solutions, reducing emissions, and enhancing fuel efficiency for traditional vehicles. Initiatives also promote the adoption of AI-driven technologies in public transport, including CCAM solutions.
- Human-centric innovation, covers the projects aimed at addressing the needs and requirements of users about the AA and AI systems under development. This research generally includes human factors, reskilling and upskilling, and user experience, both from an operator and traveller perspective.

Against this background, the analysis of the state of the art and the main SRIAs emphasised three key transversal themes, likewise technological feasibility, standardisation and certification and just transition and acceptability. Accordingly, the priorities can be mapped as follows:

- Technical and technological challenges mainly concern:
    - Data harmonisation, generally encompassing the issues concerning the quality of data – and, as a consequence, sensor technologies – and connectivity (accessibility issues will be addressed in the standardisation);
    - Infrastructure digitalisation, including all the infrastructure enhancement, development and maintenance aimed to improve compatibility with AA- and AI-based concepts;
    - Vehicles automation, empowering vehicles and functions able to enhance mobility safety and security, supporting users' needs in critical situations and reducing the number of incidents;
    - Networks management, developing AA- and AI-based systems able to optimising the transport and mobility networks capacity and efficiency;
    - Cybersecurity, ensuring the resilience of connected and automated parts and systems of vehicles and infrastructures against cyber threats.

- Standardisation and certification challenges generally address:
    - General and sectoral regulatory frameworks, providing rules to safely address the specific needs of new technologies development, testing and deployment (e.g. data governance and accessibility, risk management and protection by design);
    - Standardisation, developing shared technical and industrial standards to ensure consistency and interoperability, especially at transnational level;
    - Certification programme, fostering the design of certification processes able to effectively validate the innovative requirements of AA- and AI-based systems, ensuring trustworthiness over time

- Just transition and acceptability challenges, eventually, encompass:

- Workforce training and new job creation, ensuring adequate upskilling and reskilling programmes to smooth the transition for the workers mostly affected by the technological innovation, also creating new jobs in fully automated and connected mobility systems and services;
- Ethical and social impact assessments, promoting a proactive approach to vehicles, infrastructures and services envisioned since the early stages of design, also taking into account accessibility, fairness, social well-being and environmental impacts;
- Encourage smart urban planning and mobility, prioritising smart, sustainable and integrated mobility systems and fostering users' behavioural changes.

3.5 Takeaway messages

The development and implementation of solutions based on advanced automation and intelligence in the mobility and transport sector reveal interesting commonalities, despite the specific differences between the different transport modes. In particular, the range of opportunities in terms of safety, security and environmental impact transcends individual modes, highlighting their cross-cutting nature. In light of the above, these are the main takeaway messages:

- Safety as a priority of all the transport modes:
  - Solutions based on advanced automation and intelligence in mobility and transport share commonalities despite mode-specific differences.
  - Opportunities in safety, security, and environmental impact extend across individual transport modes, emphasising their cross-cutting nature.

- Holistic approach to sector challenges:
  - Addressing challenges in the sector requires a comprehensive approach.
  - Research and development of new technologies should consider both technical requirements and societal needs to minimise the impact on workers, users, and society.

- Transversal research and development:
  - A transversal approach to research and development is crucial for tackling sector challenges effectively.
  - Analysing main challenges underscores the importance of considering the broader implications and interconnectedness of new technologies.

- Human-centred approach to new technology regulation:
  - Policy guidelines emphasise the need for a regulatory framework tailored to the specificities of advanced automation and intelligence in transport.
  - Encouraging a holistic and interdisciplinary approach, the policy aims to strike a balance between technological progress and societal well-being.

- Balancing innovation and societal well-being:
  - The regulatory, normative, and certification framework should be adapted to the unique characteristics of advanced automation and intelligence.
  - This approach seeks to ensure that the deployment of innovations benefits both industry stakeholders and the wider community while minimising negative societal impacts.

4 Higher Automation and AI in aviation

4.1 Introduction

Automated systems and AI-supported technologies receive an increasing focus in aviation. For this domain, a comprehensive literature survey of the actual trends and advancements in these topics is provided to give an overview on state-of-the-art such as the current methods and technologies which are being planned, developed or already applied. The objective is to identify similarities and differences of the adopted solutions and the benefits and challenges placed at operational level.

For purposes of the project, this literature survey focuses on airspace optimization and human assistance in connection with higher automation as these aviation-related topics cover the four case studies that are used to support the design and the validation of the holistic and unified approach to certification. The four case studies defined by HUCAN are:

- UC1# - Dynamic Airspace Sectoring
    - Purpose: Improvement of middle airspace utilisation obtained by means of dynamic optimization of the airspace sector configuration.
    - Objective: Dynamically define and apply the best allocation of elementary sectors for the optimization of air traffic controllers (ATCOs) workload, sector capacity and flow management.

- UC2# - AI-Powered Digital Assistant in Terminal Manoeuvring Area (TMA)
    - Purpose: Enhance runway efficiency by optimising aircraft routing, ensuring adherence to procedures, and preventing potential conflicts.
    - Objectives:
        a) Assigning the quickest routes to aircraft while minimising approaching queue length and adhering to International Civil Aviation Organization (ICAO) spacing rules. This is achieved by modifying flight paths from FCFS strategy, increasing runway capacity and throughput.
        b) Maximising adherence to CDO procedures, with environmental impact reduction.
        c) Ensuring continuous CDR functionality (safety increased) by proactively identifying possible LOS, defined as simultaneous violations of horizontal distances (<5 NM) and vertical distances (<1000 ft), and taking appropriate actions to prevent them (by Reinforcement Learning technique).
        d) Workload reduction for ATCO and Pilot.
        e) Reduction of fuel consumption.

- UC3# - Dynamic Airspace Reconfiguration Service for U-Space
    - Purpose: Dynamic U-Space volumes definition and information exchanges between ATM and U-space.
    - Objective: Tool and AI Application dynamically support ATCOs in shaping, activating/deactivating U-Space volumes to UAS traffic for management of priority operations, emergencies, of manned aviation in U-Space, with benefits in optimization

of U-Space as well as controlled airspace, increase of safety levels and ATCO workload reduction.

- UC4# - Dynamic Allocation of traffic between ATCO and system
    - Purpose: Improvement of upper airspace utilisation by means of dynamic allocation of traffic between the ATCO and ATC Real Ground-breaking Operational System (ARGOS).
    - Objectives: Dynamically support the ATCOs in managing the traffic in the sector, by means of issuing operational clearances to safely handle basic traffic situations and aid controllers in handling complex traffic situations. ARGOS has 3 modes of use. Two of them will be taken into account in HUCAN: the autonomous management of the traffic by ARGOS in specific circumstances and the hybrid management of the traffic between the ATCO and the ARGOS system (dynamic allocation of traffic).

These case studies map the challenges that are in the SRIA as are particularly associated with certification issues. They cover different aspects of the capacity-on-demand concept, address different kinds of airspaces (i.e., middle airspace, TMA, U-space), and are based on different technologies and kinds of algorithms (both deterministic and non-deterministic AI-powered ones). Finally the case studies will be used to feed and validate the theoretical research, to design and test the certification method and to produce and validate guidelines for certification.

Accordingly, the scope of this literature survey is focused on airspace optimization and human assistance covering these HUCAN case studies and includes the control-centre (TMA, lower, upper airspace), the tactical phase and the ad-hoc phase (between 2 minutes to 2 hours before flight). The working places of tower and airport controllers as well as the pre-tactical planning phase (between 2 and 12 hours before flight) are not considered.

4.2 Airspace Optimization

Work on optimising airspace and associated procedures has been carried out in recent years with two main objectives. One was to improve traffic flows by adapting main flight routes or sector shapes. Secondly, the work has focussed on the integration of various aircraft systems. The focus here was on the integration of UAVs into conventional airspace.

To classify the different approaches to airspace adaptation and route optimisation, the literature researched was divided into different groups. These include Dynamic Airspace Configuration, Human-Autonomy Teaming, and Planning System Development (Figure 3).

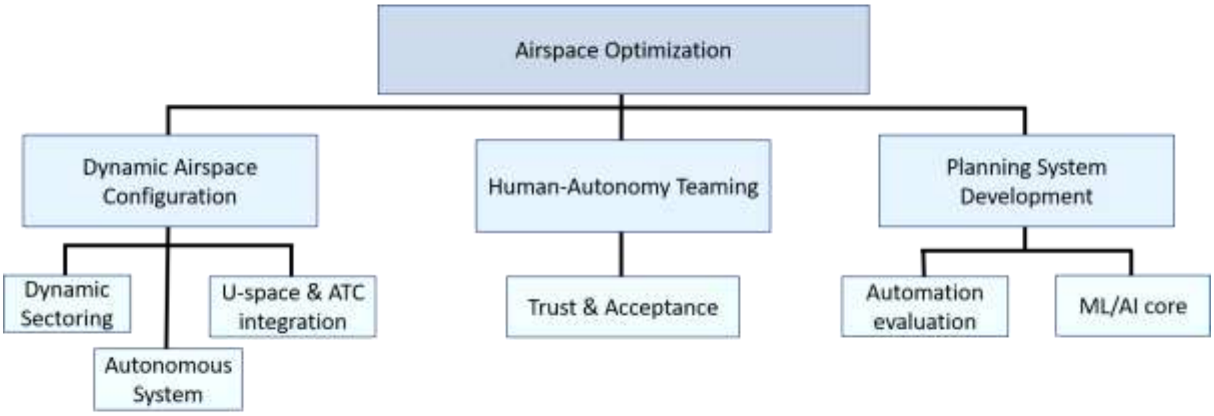


Figure 3 Classification of Literature survey of Airspace Optimization

Table 3 Overview of the literature survey, with types, subtypes and title

Classification	Sub-classification	Type	Title
Dynamic Airspace Configuration	Dynamic Sectoring	Survey	Automatic Airspace sectorization: A Survey
		Method	Dynamic airspace sectorization for flight-centric operations
			3D airspace design by evolutionary computation
		Evaluation	Validating Dynamic Sectorization for Air Traffic Control due to Climate Sensitive Areas: Designing Effective Air Traffic Control Strategies
	U-space & ATC Integration	Concept	On the Impact of UAS Contingencies on ATC Operations in Shared Airspace
			Collaborative ATM-U-space interface
	Autonomous System	Method	Optimization-Based Autonomous Air Traffic Control for Airspace Capacity Improvement
			Automated Flight Planning of High-Density Urban Air Mobility
	Trust & Acceptance		A Methodological Framework of Human-Machine Co-Evolutionary

EUROPEAN PARTNERSHIP
Co-funded by the European Union

Human-Autonomy Teaming		Method & framework	Intelligence for Decision-Making Support of ATM
			An Explainable Artificial Intelligence (xAI) Framework for Improving Trust in Automated ATM Tools
Planning System Development	ML/AI core	Methods & model development	Data-Driven Approach Using Machine Learning for Real-Time Flight Path Optimization
			A Multi-task Learning Approach for Facilitating Dynamic Airspace Sectorization
	Automation Evaluation	Evaluation	Scoring Mechanism for Automated ATC Systems

4.2.1 Dynamic Airspace Configuration

Dynamic sectorization refers to the short and medium-term adaptation of sectors to current or expected traffic volumes. The aim is to adapt the boundaries of sectors in such a way that air traffic can be managed efficiently without compromising safety. The scientific literature describes various methods of how sector boundary optimisation could work.

4.2.1.1 Automatic Airspace sectorization: A Survey (Flener et al., 2013)

In the realm of air traffic management, the paper "Automatic Airspace sectorization: A Survey", authored by Pierre Flener and Justin Pearson from the Department of Information Technology at Uppsala University, Sweden, published in 2013, stands as a pivotal exploration into the intricate world of airspace sectorization.

The survey provides a thorough examination of the concept of airspace sectorization, a critical component in air traffic management, aimed at minimising a cost metric while adhering to geometric and workload constraints. With a focus on algorithmic aspects, the paper targets experts in the field.

Distinguishing between airspace sectorization and configuration, the survey underscores the tactical nature of airspace sectorization. Configuration, described as a (pre-)tactical action, is contrasted with sectorization, which is either strategic or (pre-)tactical based on inputs. The absence of temporal aspects in sectorization, unlike configuration, presents challenges in reusing models. The paper serves as a technical overview for air traffic control (ATC) and ATM experts, emphasising algorithmic aspects and excluding realism evaluations.

The survey introduces classification criteria, categorising approaches into graph-based and region-based models. It classifies frequency as static or dynamic and explores input and output granularity, dimensionality, constraints, workload categories, constraint types, and cost functions. The technology section discusses various algorithm design methodologies and optimisation technologies, including hybrid approaches. Test scales and data types for evaluating airspace sectorization tools are also outlined.

Table 4 Airspace sectorization classification criteria provided in (Flener,P. & Pearson, J. 2013)

Name	Description
Approach	Graph-based model Region-based model
Frequency	Static: strategic or pre-tactical Dynamic: tactical at pre-determined times
Input Granularity	Mesh of blocks ATC functional Blocks (AFBs) Elementary Sectors Control sectors Area of Specialisation (AOS) Air Traffic Control Center (ATCC)
Output Granularity	functional airspace Blocks (FABs) Elementary Sectors Control sectors Area of Specialisation (AOS) Air Traffic Control Centre (ATCC)
Cost Function	Coordination cost: total cost of coordination between sectors Transition cost: cost of switching from old to new sectors Workload imbalance: impact of resulting sectors on workload balance Number of sectors: the total number of sectors should be minimised Entry points: minimise total entry points into resulting sectors
Technology	Stochastic local search (SLS) Constraint programming (CP) Mathematical modelling (MP)

	Global optimization (GO)
	Evolutionary algorithms (EA)
	Computational geometry
	Ad hoc algorithm design

Encompassing 16 approaches from 1998 to 2011, the survey provides a comprehensive overview of algorithmic methods for automatic airspace sectorization.

The conclusion stresses the need for further modelling in airspace sectorization to align with Functional Airspace Blocks (FABs) and address operational constraints. Implementing airspace optimizations that alter control sectors is acknowledged, with emphasis on the associated heavy costs in training and potential infrastructure changes in Air Traffic Control Centres (ATCCs). Transition costs are highlighted, requiring careful planning for changes in airspace design. The survey recommends increased use of constraints in computation processes, advocating for mature optimisation technologies such as Constraint Programming (CP) and Mathematical Programming (MP). The separation of concerns between modelling and solving is deemed crucial for flexible exploration in an evolving field like sectorization.

4.2.1.2 Dynamic airspace sectorization for flight-centric operations (Gerdes et al, 2018)

In the scope of air traffic management, the paper "Dynamic airspace sectorization for flight-centric operations" [1], authored 2018 by Ingred Gerdes, Annette Temme and Michael Schultz from the German Aerospace centre, Braunschweig, Germany, shows a possibility to dynamically adapt sectors to main traffic routes in order to optimise the efficiency of airspace.

The aim of the work was to create a suitable fast and efficient continuous airspace sectorization that can react to current traffic flows and efficiently support the controller even in unusual traffic situations. This approach bridges the gap between structured and unstructured airspace designs and will therefore be a fundamental key element for the efficient management of future urban airspace. The approach is so dynamic that it could also react to different traffic flows over the course of a day with a variable adjustment of the sector boundaries. The scalable approach follows the requirements of air traffic by bundling traffic patterns, identifying areas with high traffic density and providing an efficient planning and control structure to support airspace operators and users.

To ensure a more efficient allocation and a harmonised distribution of workload, the paradigm of traffic flow determined by the airspace structure ('flow follows structure') has been transformed into a dynamic approach in which the structure is adapted to the traffic flow ('structure follows flow'). The benefits of dynamic sectorization can include improved capacity utilisation through flexible use of airspace and better distribution of the workload for air traffic controllers.

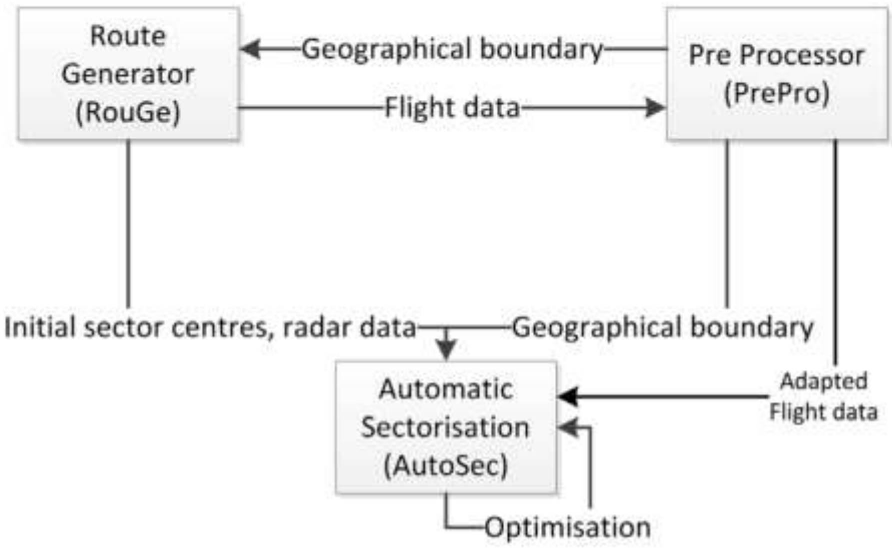


Figure 4 Generation of sector structure depending on task load in (Ingrid et al. 2018)

Different parameters for the structure of sectors are stored in a chromosome set. The main flight routes were extracted from DDR2 data sets of EUROCONTROL and summarised using a fuzzy clustering method. The airspace was initially constructed as a Voronoi diagram, which already contained the corresponding centre points.

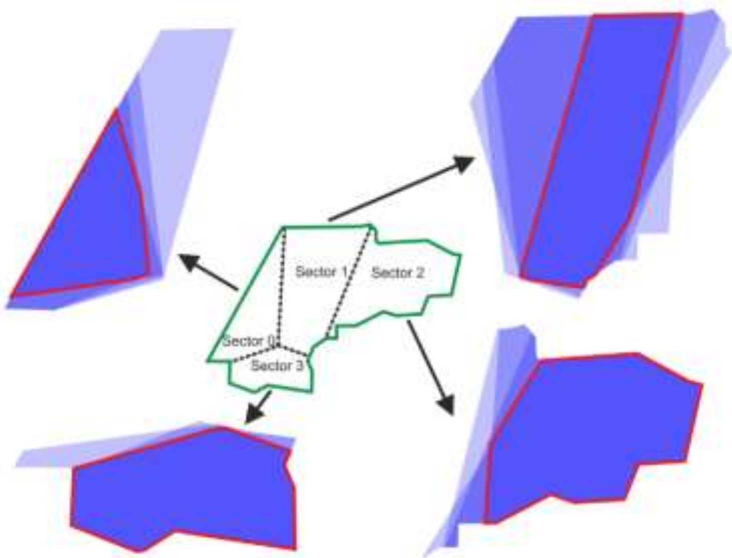


Figure 5 Example for the overlapping sectors (blue areas, the darker, the more sectors overlap) and the SBBs (bordered by red lines) (Ingrid et al. 2018)

An evolutionary algorithm then performed the optimization over several generations, creating a population with a predefined number of solutions for the given problem, where each solution is encoded as a sequence (chromosome) of parameters (genes) describing a possible problem solution. As in nature, solutions between two chromosomes can be mixed or mutated by crossover. An evaluation function decided after each generation which chromosomes were the most suitable for the next generation.

4.2.1.3 3D airspace design by evolutionary computation (Delahaye, et al. 2008)

In the scope of air traffic management, the paper "3D airspace design by evolutionary computation", authored 2008 by Daniel Delahaye and Stephane Puechmorel from Ecole Nationale de l'Aviation Civile (ENAC), Toulouse, France, the used genetic algorithm stands in the foreground, not the tactical optimization of the air space to optimise the traffic flows or controller's workload [2].

This paper from basic research presents an airspace-cutting method which synthesises balanced sectors with minimum flow cut. It shows a way to divide an airspace into meaningful sectors that must obey various boundary conditions. These include convexity, minimum distances from route crossing points to sector boundaries, minimum dwell times of aircraft in sectors and the specification of sector boundaries running vertically in space. The approach chosen to generate sectors, which must always be designed in such a way that there are no gaps between them, is an evolutionary algorithm that uses techniques of inheritance, mutation, selection and recombination (crossover) inspired by nature to find (near) optimal solutions to complex problems. The modifiable parameters are encoded on chromosomes and only evaluated after each generation using a fitness function. It was shown that even with very large airspaces and hypothetical 1000 sectors, the algorithm very quickly produces good results that meet all boundary conditions.

From an actual point of view, the results from the paper are not suitable for operational use, where en-route sectors are to be dynamically adapted to current demand. Instead, very large airspace could be created in this way according to the required criteria on the basis of a complete reorganisation. Due to formatting errors, some equations are difficult and sometimes impossible to read.

4.2.1.4 Validating Dynamic Sectorization for Air Traffic Control due to Climate Sensitive Areas: Designing Effective Air Traffic Control Strategies (Ahrenhold et al, 2023)

The study titled "Validating Dynamic Sectorization for Air Traffic Control due to Climate Sensitive Areas: Designing Effective Air Traffic Control Strategies", authored by Nils Ahrenhold, Ingrid Gerdes, Thorsten Mühlhausen, and Annette Temme from the German Aerospace Centre (DLR) Braunschweig, Institute of Flight Guidance, explores the application of dynamic sectorization in air traffic control to address challenges posed by climate-sensitive areas. Published in 2023, the research focuses on validating the effectiveness of dynamic sectorization strategies, aiming to balance the workload of air traffic controllers amid changing traffic patterns influenced by climate-related considerations. This summary provides an overview of the key findings, methodologies employed, and implications for enhancing air traffic management in response to dynamic environmental factors.

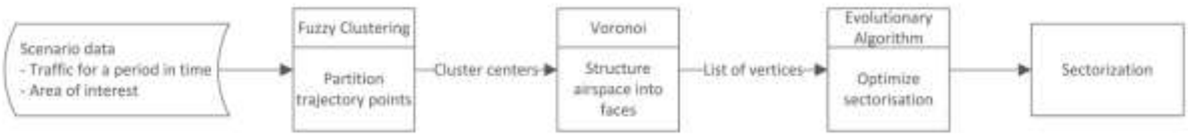


Figure 6 Overview of the dynamic sectorization approach taken (Ahrenhold, et. al 2023)

In this feasibility study, the application of dynamic sectorization in the context of air traffic control to cope with climate-sensitive areas is examined. Dynamic sectorization serves as a means to balance the workload of air traffic controllers in response to changing traffic patterns. A multi-objective optimization system analyses traffic flow and determines time-dependent sectorizations, integrated into a radar display. The use case involves climate-sensitive areas causing changes in traffic flow.

The study evaluates the developed solution through Human-in-the-Loop (HITL) tests with air traffic controllers. A controller assistance system in a dynamic airspace sectorization environment is compared with traditional working methods. The validation shows that the solution is highly applicable according to controllers' assessments, yet emphasising the need to adapt current procedures and define new aspects more precisely.

The methodology includes the application of a three-stage approach for dynamic sectorization, incorporating Fuzzy Clustering, Voronoi Diagram, and Evolutionary Algorithms. Two defined scenarios are used to test the system's performance, including a climate-sensitive scenario with changes in traffic flow due to contrail restrictions.

The results of HITL experiments show that the DAS approach is effective without unrealistic behaviours in the simulation. Air traffic controllers rate the system's performance as realistic, with no safety concerns or increased workload. Suggestions for improving controller guidelines are made, including naming conventions and visual guidelines.

In the conclusion, it is emphasised that the study represents an initial feasibility study and clear responsibilities are necessary for sector adjustments. Recommendations for dealing with specific situations and initial guidelines for using dynamic sectorization are developed. Future steps include adjusting the evaluation function based on controller feedback and extended feasibility tests with interacting controllers. The method could also be expanded to the 3D airspace to enable horizontal and vertical sector management.

4.2.1.5 On the Impact of UAS Contingencies on ATC Operations in Shared Airspace (Teutsch, et al. 2023)

In the publication "On the Impact of UAS Contingencies on ATC Operations in Shared Airspace" from March 2023, the authors J. Teutsch, C. Petersen, G. Schwoch, T. J. Lieb, T. Bos and R. Zon share findings of simulations for the SESAR Industrial Research Project AURA, which were carried out by the Royal Netherlands Aerospace Centre, NLR, together with partners from the German Aerospace Centre, DLR.

It is expected that these new airspace users will extend their operations and share available airspace with manned traffic. Dynamic Airspace Re-configuration (DAR) has been considered as one of the enablers for the integration of unmanned and manned traffic in such non-segregated airspace.

AURA investigates requirements for an interface between ATM-controlled airspace and highly automated U-space airspace for large numbers of unmanned aircraft. The project defined AUSA, an ATM U-space Shared Airspace, which is a generic type of airspace that can be delegated to contain both ATC and U-space controlled airspace volumes and identified the flow of information between actors, roles and services (ANSP, CISP: Common Information Service Provider, USSP: U-space Service Providers). Operational Environment, Human Performance Challenges and DAR are described in this paper.

Figure 7 Surveillance Display used for DAR Manager and ATCO - Active (Solid) and Planned DAR (Dashed) (Teutsch, et. al

The AURA concept follows the principle to be in line with existing research activities and regulatory framework developments in Europe and is set between phase U3 (U-space advanced services) and U4 (U-space full services) of the SJU.

Results of described simulations:

- An introduced DAR Manager role and the designed working position supported and improved ATC operations.
- Negotiations between the DAR Manager and air traffic controllers, will only be possible if there is enough lead time (several minutes) to prepare for airspace changes.
- Emergency requests that require immediate action should be communicated to the affected controllers immediately by the system.

4.2.1.6 Collaborative ATM-U-space interface (López et al, 2023)

In the preprint “Collaborative ATM-U-space interface” from October 2023, the authors M. M. López, M. C. Gutiérrez published a concept within the AURO project led by Indra. By developing a concept of operations and validating an identified set of selected information-exchanges services between ATM and U-space systems by identifying the requirements for USpace information exchange with ATM through SWIM, the foundations were laid for the integration of the new entrants in the current and future air traffic environment.

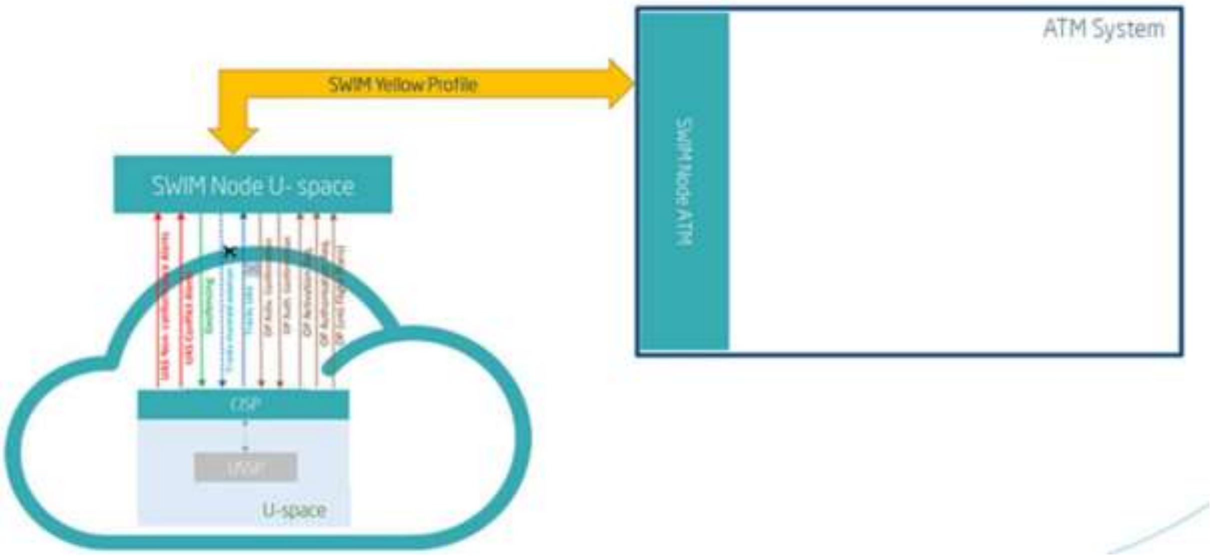


Figure 8 Information exchange services architecture from between ATM and U-space (López, et al. 2023)

The proposed solution focused on the generation of a common ATM-U-space interface by identifying an initial set of basic services considering the relevant information needed to be exchanged so as to permit and guarantee the interoperability between both systems, avoiding airspace fragmentation and allowing safe drones' operations into controlled airspace. The exchange shall ensure the necessary information is available to the related stakeholders in order to enable coexistence of ATM and U-space traffic.

The Assessment of Validation Objectives regarding U-space operations in controlled airspace takes into account:

- Operational acceptability of roles, tasks and operations.
- Technical feasibility of support.
- Suitability of the ATM-U-space interface for the different solution architectures.
- Impact on human performance.
- Impact on overall safety of U-space operations in controlled airspace.
- Different operating concepts in terms of missions, operational procedures, information exchanges and architecture configurations.

The findings of the validation regarding airspace optimization are:

- Information exchanges between ATM and U-Space (and vice versa) for sharing new volumes definition (dynamic U-space Airspace reconfiguration means U-space airspace volumes modifications).
- Findings regarding human performance challenges:
- The controller workload was hardly impacted by the activity of this interface. Tasks can be performed efficiently and safely.
- Situational awareness remains at high levels, but HMI needs to be improved to minimise controller interventions.

4.2.1.7 Optimization-Based Autonomous Air Traffic Control for Airspace Capacity Improvement (Başpınar et al, 2020)

In the 2020 publication, "Optimization-Based Autonomous Air Traffic Control for Airspace Capacity Improvement", Barış Başpınar and Hamsa Balakrishnan from the Massachusetts Institute of Technology in Cambridge, USA, collaborated with Emre Koyuncu from Istanbul Technical University, Turkey. The paper responds to the increasing demand in air traffic by introducing an innovative autonomous ATC system rooted in optimization. The study aims to cope with the rising demand in air travel through highly automated assistance.

The core of the paper lies in introducing an optimization-based autonomous ATC system with a specific focus on determining airspace capacity. The study highlights the critical role of predicted trajectories in the decision-making process and underscores the significance of simulating aircraft movements to estimate airspace capacity accurately.

To achieve accurate trajectory predictions, the paper model's aircraft dynamics and guidance procedures. These models simulate aircraft movements, contributing to the overall predictive capability of the proposed ATC system.

Predicted trajectories emerge as pivotal components influencing decision-making, and the simulation of aircraft movements is crucial for creating a traffic environment conducive to estimating airspace capacity accurately.

The interventions of an air traffic controller are defined as a set of manoeuvres suitable for real air traffic operations, providing a human-compatible touch to the autonomous system.

The decision-making process of the designed ATC system relies on Integer Linear Programming (ILP). ILP is constructed through a mapping process, discretizing airspace with predicted trajectories and enhancing the temporal performance of conflict detection and resolution.

The paper introduces a method for estimating airspace capacity using the proposed ATC system. The procedure involves constructing a stochastic traffic simulation reflecting the structure of the evaluated airspace.

Validation of the approach is conducted using real air traffic data for en-route airspace, ensuring the practical applicability and reliability of the proposed ATC system.

The study concludes by showcasing the effectiveness of the designed ATC system in managing air traffic, even under higher density conditions than current air traffic scenarios. It also concludes by acknowledging that the proposed system, though depicted as fully autonomous, can also function as a semi-autonomous system for decision support by human air traffic controllers. The decision on the autonomy level rests with authorities, who can choose based on stakeholder preferences and other factors. The scalability and applications of the system are highlighted, with the ILP formulation enabling scalability for large-scale ATM scenarios. The benefits and drawbacks are discussed, emphasising high scalability, easy integration into existing ATM systems, and the use of realistic models to avoid operational hazards. Future research directions are suggested, focusing on expanding the method to handle different multi-agent systems and exploring alternative complexity metrics, constructing a detailed wind model, and improving the aircraft model for more accurate trajectory predictions.

4.2.1.8 Automated Flight Planning of High-Density Urban Air Mobility (Tang et al, 2021)

In the realm of advancing urban air mobility, the paper "Automated Flight Planning of High-Density Urban Air Mobility", authored by Hualong Tang and Yu Zhang from the Department of Civil and Environmental Engineering, alongside Vahid Mohmoodian and Hadi Charkhgard from the Department of Industry and Management Science Engineering, all affiliated with the University of South Florida, USA, presents a pioneering exploration into the intricacies of automated flight planning systems. Published in 2021, this research delves into the challenges and requirements posed by the burgeoning field of high-density urban air mobility, aiming to provide scalable, safe, and autonomous solutions.

The study proposes an Automated Flight Planning System (AFPS) to address the anticipated higher density of AAM operations. The AFPS components, including the Low-Altitude Airspace Management System (LAMS) and Low-Altitude Traffic Management System (LTMS), aim to provide scalable, safe, and autonomous solutions.

To meet the demands of high-density operations, the paper recommends third-party service providers for air traffic management and introduces the AFPS. The components of AFPS involve innovative technologies like LiDAR data for 3D map generation and the visibility graph method for nodal network construction.

The LTMS focuses on designing conflict-free 4D trajectories based on flight requests, considering system cost and equity among operators. The Nash Social Welfare Program (NSWP) is introduced to maintain fairness among different operators in case of service provided to multiple UAM operators.

A case study in the Tampa Bay area in Florida serves to demonstrate the operability of AFPS, showcasing conflict-free UAM operations through animations. The paper also discusses tactical operational decisions for electric vertical takeoff and landing (eVTOL) vehicles, emphasising a shift from traditional flight planning.

The literature review outlines current challenges in airspace design, the UAM corridor concept proposed by FAA, the visibility graph method, conflict detection methods, trajectory deconfliction approaches, and the importance of flight equity in U-space services. The proposed AFPS is presented as a solution, aiming to generate nodal networks that avoid obstacles in low-altitude urban airspace.

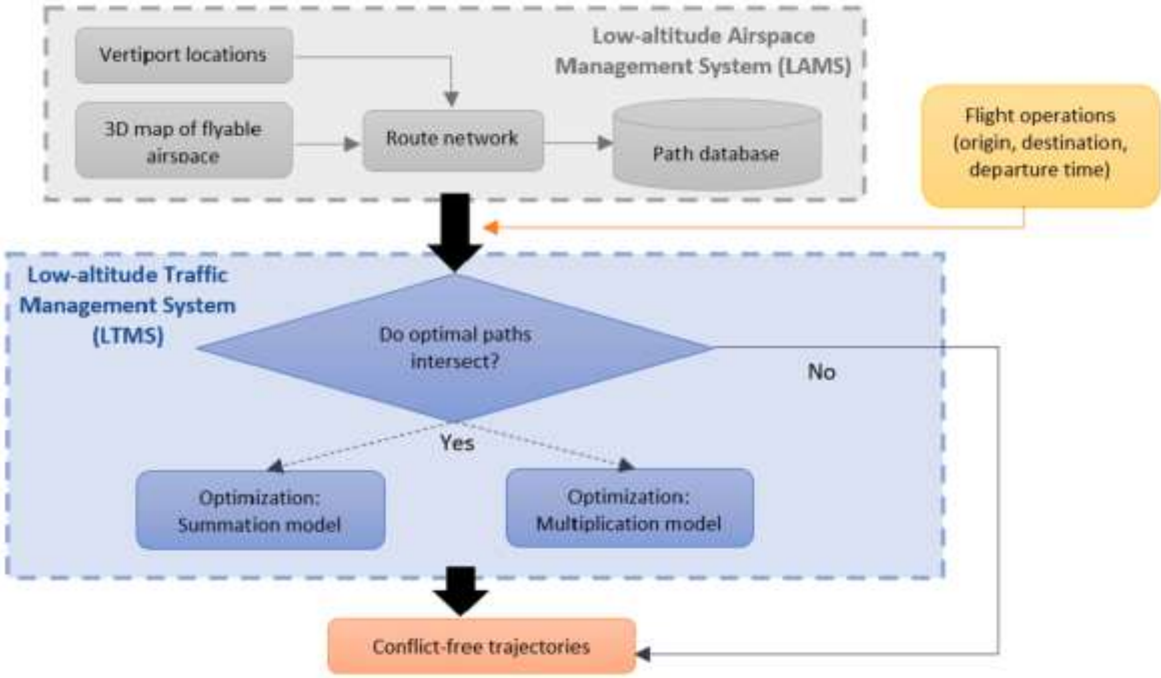


Figure 9 Workflow described (Tang, et al. 2021)

Experiment results include the comparison of models and solvers, analysis of optimality and trajectories, model comparison, flying time analysis, solution differences, system costs, Operators' Unfair Benefit Ratios (UBRs), and computation time. The conclusion emphasises the development and components of AFPS, the importance of fairness principles, the LTMS fairness demonstration, the success of the case study, and ongoing research and future directions.

The study acknowledges ongoing research areas, including additional conflict resolution strategies, integration of strategic and tactical planning, applicability to other types of AAM, modification of cost functions, consideration of weather patterns, and integration of battery monitoring into operational decisions.

4.2.2 Human Autonomy Teaming

Human Autonomy Teaming (HAT) refers to the interaction of people with automatic or semi-automatic systems (Lyons et al. 2021). Depending on the activity and trust in the systems, the acceptance of people in particular is a challenge for support system development in a professional and safety-critical environment.

4.2.2.1 A Methodological Framework of Human-Machine Co-Evolutionary Intelligence for Decision-Making Support of ATM (Hu, 2020)

The study "A Methodological Framework of Human-Machine Co-Evolutionary Intelligence for Decision-Making Support of ATMs", authored by X. B. Hu proposes a methodological framework of human-machine co-evolutionary intelligence (HMCEI) for decision making support of ATM. As long as an AI method aims to compete and replace human controllers, it will be confronted with the difficulty of not being accepted by human controllers. To address this dilemma, this paper proposes a new thinking about applying AI methods, i.e., an AI method should be developed in such a way to assist human controllers, but never in the way of competing and replacing human controllers.

Although the study is about any implementation, the proposed approach to make artificial intelligent (AI) methods more acceptable in ATM might be an enabler for LOAT level 2.

4.2.2.2 An Explainable Artificial Intelligence (XAI) Framework for Improving Trust in Automated ATM Tools (Hernandez et al, 2021)

The paper "An Explainable Artificial Intelligence (XAI) Framework for Improving Trust in Automated ATM Tools" by C. S. Hernandez, S. Ayo and D. Panagiotakopoulos, describes the basis of an XAI Trust framework in order to address the gap between research and implementation solutions within an ATM environment. It highlights current guidelines and recommendations by regulators for trustworthy AI and addresses what constitutes trust in AI automated solutions in ATM for end users through an AI Trust Survey answered by stakeholders of the Fly2Plan project.

4.2.3 Planning System Development

Tactical and pre-tactical planning systems for air traffic controllers have been developed for over thirty years. Arrival (AMAN) and Departure Managers (DMAN), for example, are in use at many international airports and can now be purchased commercially and customised for the respective airports. For upper airspace, there are en-route managers that support the organisation of airspace. While these systems were previously based on deterministic algorithms, initial attempts have been made in recent years to develop AI-based air traffic controller support systems. The aim is to simulate controller behaviour more realistically in different situations and thus also improve the HAT.

4.2.3.1 Data-Driven Approach Using Machine Learning for Real-Time Flight Path Optimization (Kim et al, 2022)

The pursuit of efficient in-flight replanning amidst changing weather conditions has led to the development of an automated framework explored in the paper titled "Data-Driven Approach Using Machine Learning for Real-Time Flight Path Optimization", authored by Junghyun Kim, Cedric Justin and Dimitri Mavris from the Georgia Institute of Technology, Atlanta, Georgia, along with Simon Briceno from Jaunt Air Mobility, Atlanta, Georgia, published in 2022. This study addresses the challenges faced by airlines due to flight delays caused by convective weather. The study sets out to create an automated solution leveraging supervised and unsupervised machine learning techniques along with a graph-based pathfinding algorithm. The primary objective is to minimise operational costs for airlines by generating optimised flight paths.

The challenges of manual in-flight replanning and existing limitations in solutions like NASA's Traffic Aware Planner prompt the need for an advanced approach. The study advocates for the integration of AI to enhance flight planning, filling the gap in real-time weather optimization. The proposed automated framework utilises supervised machine learning for wind regression, unsupervised machine learning for short-term convective weather forecasting, and optimised flight path generation based on designated points.

The paper provides a comprehensive overview of the proposed methodology, which involves a data-driven approach for precise and frequent in-flight replanning. Leveraging supervised machine learning, a hybrid algorithm for wind modelling is introduced, encompassing Multilayer Perceptron (MLP), Support Vector Regression (SVR), and Gaussian Process (GP) techniques. Unsupervised machine learning techniques, specifically DBSCAN, are employed for short-term convective weather modelling.

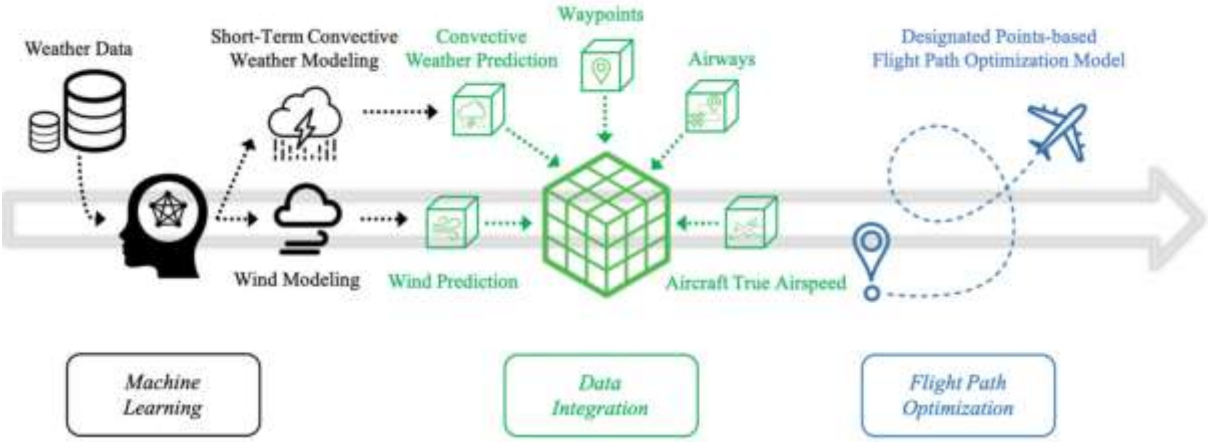


Figure 10 Methodological overview of the proposed solution (Kim, et. al 2022)

Flight path optimization, based on designated points, incorporates a hybrid method combining the A* search algorithm with a Free-Flight approach. The approach addresses various assumptions, such as constant aircraft speed during the en-route phase and the representation of convective weather by polygons incurring high penalty costs when penetrated by an aircraft.

Two comprehensive case studies on Delta Airlines flights under varying weather conditions validate the effectiveness of the proposed framework. The first case study demonstrates a minimal difference between the real and simulated flight times, indicating the reliability of the developed system. The second case study, conducted during heavy weather, reveals a significant reduction in simulated flight time, showcasing the potential benefits of the framework in adverse conditions.

The study concludes by presenting three distinct approaches to flight path optimization, all rooted in machine learning for wind regression, weather forecasting, and path optimization at designated points. The statistical analysis of real flight data emphasises that the proposed framework consistently generates flight routes reducing flight time by up to 2%. The developed system empowers US airlines to conduct more accurate and frequent flight path optimizations, with further opportunities for improvement highlighted through the integration of additional operational constraints.

4.2.3.2 A Multi-task Learning Approach for Facilitating Dynamic Airspace Sectorization (Zhou et al, 2022)

The publication “A Multi-task Learning Approach for Facilitating Dynamic Airspace Sectorization” authored by W. Zhou, Q. Cai, S. Alam proposes a multi-task learning (MTL) approach which is able to predict sector traffic flow and airspace capacity simultaneously using a shared neural network architecture. Specifically, the proposed model predicts the demand-capacity imbalance and identifies the opportunity for sector split/merge implementation. This method is a promising approach but needs to be validated.

4.2.3.3 Scoring Mechanism for Automated ATC Systems (Movila, 2023)

In the scope of air traffic management, the Bachelor thesis "Scoring Mechanism for Automated ATC Systems", authored 2023 by George-Daniel Movilă from Escola d'Enginyeria de Telecomunicació i Aeroespacial de Castelldefels, Universitat Politècnica de Catalunya, Barcelona, Spain, describes a scoring system for quantitative assessment of an upper airspace controller support system. The scoring mechanism focuses on providing a scoring function that evaluates both the operational safety and the efficiency of the trajectories proposed by MUAC's ARGOS (EUROCONTROL 2023).

MUAC's ARGOS system is designed to improve the situational awareness of air traffic controllers, reduce their workload and enable an increase in capacity in certain situations. The scoring mechanism of this thesis focused on providing a scoring function that evaluates both the operational safety (Decreasing Score, D-Score) and the efficiency of the trajectories proposed by ARGOS (Increasing Score, I-Score). In the author's opinion, automatic scoring should always be used in aviation where decisions are made automatically. However, the proposed evaluation system only works offline on the basis of log files from simulations.

The scoring mechanism developed can contribute directly to the automation of air traffic control. Firstly, the mechanism can ensure that the automated air traffic control systems make correct decisions, as the algorithm evaluates the software according to how closely its decisions match human decision-making standards. This reduces the likelihood of errors and increases safety. Furthermore, the scoring mechanism provides a way to also assess the performance of automated systems and evaluate the system's ability to handle different types of traffic situations. Finally, the developed scoring provides continuous feedback that can help the system improve its performance after each update. Finally, specific metrics allow detailed insights into the performance of the safety-critical air traffic control systems.

A disadvantage of the scoring mechanism is that no upper limits are available for certain metrics. These limits could define whether the deviation from the optimal scenario is still acceptable or should be discarded. As an example, an extended scoring could identify any situation in which an aircraft deviates significantly from its planned course and should be checked by the automatic system or a controller. By adjusting the weighting of existing metrics, or adjusting them directly, it could be achieved that they better reflect the overall performance of the system.

The thesis suggests that scoring mechanisms should not only take data from one system, but also use information from other independent sensors or sources where possible. The paper uses examples to show where and how ARGOS could be modified to improve the informative value of scoring. The scoring calculation includes the following 17 parameters:

1. Violation of separation standards (separation infringement)
2. Penetration of a Temporary Segregated Area
3. Trigger of a short-term conflict alarm
4. Not respecting horizontal safety buffers after HDG
5. Not respecting vertical safety buffers after reaching CFL
6. Not respecting vertical safety buffers prior to reaching CFL
7. Flights outside the controlled airspace
8. Flights that exit through other points than specified in flight plan
9. Too frequent clearances given to a flight

10. Turns too high used to solve conflicts
11. Flown more than 5% above the shortest distance
12. More than 3% of the flown distance prior to Top of Descend not cleared to the cruising level (ECL)
13. Horizontal deviation from the exit point
14. Vertical deviation from the exit point
15. Instructions given to other points than the ones specified in the flightplan
16. Flights that do not reach the planned cruising level
17. Flights arriving too early at the transfer flight level

Most parameters are not calculated, but contain a fixed value that is added to or subtracted from the score when the situation in question occurs.

From the author's point of view, the objectives of the project were achieved. It was shown that even a simple scoring algorithm can be implemented for a complex system and thus its performance can be tested. The scoring mechanism is not limited to ARGOS from MUAC, as other automated systems such as the Advanced Autoplanner (AAP) or Skyler (an artificial intelligent air traffic controller agent) could also use it for validation.

The scoring mechanism focuses on providing a scoring function that evaluates both the operational safety and the efficiency of the trajectories proposed by ARGOS. In the author's opinion, automatic scoring should always be used in aviation where decisions are made automatically. The mechanism can ensure that the automated air traffic control systems make correct decisions, as the algorithm evaluates the software according to how closely its decisions match human decision-making standards. Furthermore, the scoring mechanism provides a way to also assess the performance of automated systems and evaluate the system's ability to handle different types of traffic situations.

4.3 Assistant to Human

Assistant to humans in aviation has progressed along the path of automation to support human in performing complex tasks. In recent years, the Human Assistant or digital assistants, with the increasing effectiveness of artificial intelligence, and other new automation technologies have been foreseen as the next expected steps for adoption across the aviation sector. From IR facial recognition and fever detector AI thermal cameras at airports, new technologies and digital assistants are increasingly expected to help streamline processes and assist in safety and efficiency improvements across aviation. Intelligent monitoring and assistance supporting the safety critical role of pilots on the flight deck, and aiding ATC to enable greater capacity and more efficient flight paths are also being suggested.

The State of the art on Digital Assistants (DA), aside from the specific current development reported in section 3.2.2 in the aviation domain, is a very extended topic including different backgrounds and different expertise and domains of knowledge. DAs include different dimensions such as trust, transparency, reliability, and the interaction with humans that can be considered transversal to any domain. Additionally, considering elements from different research domains can trigger a cross-fertilisation amongst the different sectors, and support the definition of a rigorous framework of definitions, attributes, dimensions of DAs in general and this can ease the standardisation, and also the benchmark amongst the different applications. The extension of the research topics underpinned implies a question: Which are the most relevant aspects for the state of the art on Human Assistant?

Wondering about this, the present work starts with observing the variety of terms that can be related to the Human Assistant. It provides a theoretical framework from the literature supporting a potential classification of the research and of the applications. Then it analyses the dimensions of man-machine interaction and human-AI teaming. Considering the previous analysis of the literature, what comes up is that aside from the interesting application of AI techniques that can be more and more challenging, to enable DAs, at least the following aspects should be investigated: Task Division and Allocation, Collaboration and Cooperation, Elements of trust, Explainability, and Performance Measures.

The current status of DA in Aviation is represented by different recent projects (listed in section 3.2.2). These conclude that in providing proof of concepts of DA in the flight deck, in the control tower, at the airport and in Advanced Air mobility scenarios, DA will address specific research questions, from Understanding the effectiveness of Artificial Intelligence in performing specific tasks, to investigating Explainability, or AI Design Assurance or Human-AI teaming dimensions.

DAs are based on AI and design assurance must be studied for the different techniques. Data management is part of AI and therefore also part of DAs. On the other hand, the impact of DAs on humans, as well as the benefit that can be derived from their adoption and the costs, including their impact on safety culture, have to be investigated. The question is if we are facing simple new tools or if we are looking at a revolution in the sector.

4.3.1 Human Assistant

The development of a good Taxonomy enables the definition of a comprehensive state of the art. It is a basic step that allows us to understand the completeness and the accuracy of the work. In the next sections, the lack of taxonomies explains how the domain can be very extended, explaining the issues of coverage and providing a justification for the identified research topics to understand the current status and to address opportunities and challenges. Indeed, in spite of defining the funding elements of the state of the art, the research on taxonomies can be by itself part of the state of the art to set the basis for the next work in the HUCAN project where a referenced taxonomy enables a sound work.

Human Assistance currently takes many forms and adopts in both industry and academic arenas a variety of category names, using some combination of “automated”, “digital,” “smart,” “intelligent”, “personal,” “agent,” and “assistant” (Grochow, 2020).
Generally speaking, the “assistant” can be digital tools, software agents, chatbots taking the form of robots, or simply having an interface on a computer. There isn’t a general consensus on a comprehensive taxonomy, due also to the fact that different domains are involved with different heritages.

A *digital tool* is intended as something that can automatically perform specific tasks and that can be based or not on Artificial Intelligence.

In (Sánchez, 1997) *“An **agent** is a software process that acts on a user’s behalf, performs particular functions autonomously and realises goals. An agent is versatile in changing environments and works in a team. Members of that team have complementary specialists or duplicates”* intending it as an autonomous software entity that can interact with its environment.

The Agent concept goes beyond the digital tool supporting a specific human task and looks at Artificial Intelligence to support the human in a dynamic environment. Agents in Artificial Intelligence can be categorised into different types based on how agent's actions affect their perceived intelligence and capabilities, such as: Simple reflex agents; Model-based agents; Goal-based agents; Utility-based agents; Learning agents and Hierarchical agents.

Chatbots typically interact with users via text, though images are now a common feature of Chatbot interactions and there are a number of 'bots with speech capabilities. This conversational capacity has been key to their success, as it reflects a consumer trend – the move away from voice-based channels and the embracing of chat-based channels.
Whereas a Chatbot focuses on a relatively narrow range of issues, a Digital Agent could be asked to do anything.
The literature defines a **Digital Assistant** as a concept that includes Artificial Intelligence, and goes beyond tools based on machine learning algorithms that provide data and information to the human operator. Instead, it's more like a colleague that interacts and "converses" with its human counterpart. This introduces the idea of Human-Artificial Intelligence teaming.

Such a definition adds a new perspective to the level of assistance that on the one hand is reflected by the taxonomies of automation and autonomy and on the other hand overlooks collaboration and cooperation and classifies the object with respect to the interaction with humans.

The literature provides a great variety of taxonomies on the basis of different dimensions. In (Grochow, 2020) an extensive review is proposed. The taxonomies are defined on the basis of:

- The task content of human activity that is used for assistant design instead of identifying classification criteria.
- The technology and features of design such as communications mode, direction of interaction, adaptivity, and embodiment (virtual character, voice), and so forth.
- The degree of the perceived intelligence and capability such as simple reflex agents, model-based reflex agents, goal-based agents, utility-based agents, and learning agents.
- The end-user view of "work output," and while this approach uses a somewhat subjective measurement scale, the intention is to extend the work incorporating objectives (see Figure 1)

Level 1: Work-output based on passively performing specifically assigned tasks;
Level 2: Work-output based on actively performing assigned tasks, developing related sub-tasks;

Level 3: Work-output based on using basic general knowledge and experience to understand specifically assigned duties and perform readily discerned tasks;

Level 4: Work-output based on using broad knowledge and experience, general and in the task area, to understand broadly assigned duties and perform implied tasks; and
Level 5: Work-output drawing on all available knowledge and experience from a variety of sources, general and in the task area, to infer useful duties, executed without supervision—just like you would have done them if you had the time, or even better!

Figure 11 The levels of Assistant on the basis of the output (Grochow, 2020).

Finally, robots considered as assistants are well classified. According to the degree of their intelligence, robots can be divided into two categories: functional robots and intelligent robots. Before the advent of intelligent robots, robots were primarily referred to as functional robots, whose main purpose was to perform actions that humans would not want to do or cannot do on their own. They were treated as tools to improve work efficiency and emancipate humans from manual labour and simple mental labour. Intelligent robots were invented to meet the demands of human intelligence, such as intelligence quotient (IQ) and emotional quotient (EQ). They are on turn divided in cognitive robots, understanding robots, interactive robots, and autonomous robots (Ren at Al., 2020)

In spite of a generally agreed taxonomy on the human assistant, it is relevant to note that the human assistant is also defined in terms of the human machine interaction that may play a huge role in the success of the assistant.

Humans interact with computers in many ways, and the interface between the two is crucial to facilitating this interaction. HCI is also sometimes termed human–machine interaction (HMI), man-machine interaction (MMI) or computer-human interaction (CHI) (Bansal et al., 2018), but the literature offers other ways reported in the table below.

Table 5 Different terms for man machine interaction

Term	Acronym
Man Machine Interaction	MMI
Human Machine Interaction	HMI
Human Agent Interaction	HAI
Human Computer Interaction	HCI
Computer Human Interaction	CHI
Human Machine Collaboration	HMC
Human Machine Cooperation	HMC
Human Machine Teaming	HMT

Man–machine interaction is described as an interaction and communication between human users and machines in a dynamic environment through several interfaces. Ever since humans started to build tools, there was the interaction between the humans and the machines. This interaction has evolved over time. Initially, before the Second World War, people were adjusted to fit machines. In other words, humans were trained to use the machines. However, in the Second World War, new equipment was developed so quickly that it was hard to sufficiently train humans. Therefore, the need for a systematic analysis and synthesis of the interaction between humans and machines arose. The history of Human–machine interaction can be split up into four time zones. First, in the years 1940 to 1955, developers tried to find the limits of human possibilities. New equipment was designed such that human controllers would just be able to deal with it. From 1955 to 1970, things advanced. At this time, researchers tried to model humans like machines and design products accordingly. Around 1970, electronics were advancing. Then, from 1970 to 1985, this technology was used to automate many tasks, which normally required humans. The human ended up being the controller and began becoming the supervisor. This has advanced much more since 1985 (Krupitzer et al., 2020).

HCI was first used in 1976, and it was popularised by the book, The Psychology of Human Computer Interaction published in 1983. In 1992, a HCI curriculum was developed by Hewett and other leading

Co-funded by the European Union

HCI educators to serve the needs of the HCI community. In CES 2008, Bill Gates emphasised the role of natural user interface and predicted that the way in which HCI will bring a radical change in the next few years. Thereafter, HCI researchers expounded the definition of a natural HCI by employing different approaches (Bansal et al., 2018).

The Association for Computing Machinery (ACM) defines human–computer interaction as "a discipline that is concerned with the design, evaluation, and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them". A key aspect of HCI is user satisfaction, also referred to as End-User Computing Satisfaction. It goes on to say: "Because human–computer interaction studies a human and a machine in communication, it draws from supporting knowledge on both the machine and the human side. On the machine side, techniques in computer graphics, operating systems, programming languages, and development environments are relevant. On the human side, communication theory, graphic and industrial design disciplines, linguistics, social sciences, cognitive psychology, social psychology, and human factors such as computer user satisfaction are relevant (Grochow, 2020).

Desktop applications, internet browsers, handheld computers, and computer kiosks make use of the prevalent graphical user interfaces (GUI) of today. Voice user interfaces (VUI) are used for speech recognition and synthesising systems, and the emerging multi-modal and Graphical user interfaces (GUI) allow humans to engage with embodied character agents in a way that cannot be achieved with other interface paradigms. An extensive literature providing the current status of research in such fields is provided in (Bansal et al., 2018), (Ren at Al., 2020) (Krupitzer et al., 2020)

The process of HMI can be divided into the following four steps according to the collection, transmission and analysis of data: (1) the sensor collects the environment and input signals, (2) the signal is converted into data, (3) the data is transmitted to the processing centre (4) interaction and collaboration.

As stated in (Damacharla et al., 2018) HMT can be defined as a combination of cognitive, computer, and data sciences; embedded systems; phenomenology; psychology; robotics; sociology and social psychology; speech-language pathology; and visualisation, aimed at maximising team performance in critical missions where a human and machine are sharing a common set of goals.

(Salas et al., 1992) define a team as "… an interdependent group of members, each with their own roles and responsibilities, that come together to address a particular goal "

Chapter 2 of (National Academies of Sciences, Engineering, and Medicine, 2022) presents relevant perspectives on human-AI teaming as a step beyond human-AI interaction. Teams are created to perform a variety of tasks that require the coordination of multiple interdependent individuals, and this definition does not require all team members to be human. Further, the performance of a team is not decomposable to, or an aggregation of, individual performances. This description emphasises the interdependence of team members.

A human-AI team is defined as "one or more people and one or more AI systems requiring collaboration and coordination to achieve successful task completion" (Cuevas et al., 2007). AI systems may play a variety of roles, ranging from decision-support tools to assistants, collaborators, coaches, trainers, or mediators. Within the human-AI teaming the human has to be in charge of the team, for reasons that are both ethical and practical. Not only are humans legally and morally responsible and accountable for their actions, but they also function more effectively when their level of engagement is high. While it is assumed that human-AI teams will be more effective than either humans or AI systems operating alone, in the committee's judgement this will not be the case unless humans can (1) understand and predict the behaviours of the AI system; (2) develop appropriate trust relationships with the AI system; (3) make accurate decisions based on input from the AI system; and (4) exert control over the AI system in a timely and appropriate manner.

Chapter 3 of (National Academies of Sciences, Engineering, and Medicine, 2022) addresses the most relevant dimensions of human-machine teaming: mental models, communication and coordination, and social intelligence.

Mental models are "*mechanisms whereby humans are able to generate descriptions of system purpose and form, explanation of system functioning and observed system states, and predictions of future states".*
A shared mental model is a consistent understanding and representation, across teammates, of how systems work (i.e., the degree of agreement of one or more mental models). A shared mental model includes models of the technology and equipment, models of taskwork, models of teamwork, and models of teammates.

A team mental model is a mental model of one's teammate(s) that provides an understanding of teammates' capabilities, limitations, current goals and needs, and current and future performance. Shared mental models within teams also contribute to the development of shared situation awareness.

Communication and coordination are essential for teamwork, given teamwork's interdependent nature. Team cognition can in fact be characterised as communication and coordination processes in addition to knowledge or shared models because team cognition involves more than just knowledge.

Communication can be verbal or nonverbal and can take place through various modalities, such as voice or text. Much progress has been made toward the creation of AI that understands natural human language; however, natural language processing remains a challenge for human-AI teaming. Moreover, natural language, with all its ambiguities, may not be the language of choice for effective teaming.

Communicating in a common language is just one requirement for effective teamwork. Communication also needs to be accurate and directed to the right team member at the right time or, in other words, coordinated. Effective teamwork requires "orchestrating the sequence and timing of interdependent actions".

Human teammates can make use of social intelligence for effective teaming. They can understand the beliefs, desires, and intentions of fellow teammates by developing a theory of mind (i.e., by observing their teammates' behaviours and ascribing mental states to them). There have been recent efforts directed toward providing AI with social intelligence such as the Defense Advanced Research Project Agency's (DARPA) ASIST program, for example, though this may resemble a theory of behaviour more than a full theory of mind (Sandberg, 2021).

4.3.1.1 Collaboration and cooperation

Among the many types of interactions that can take place between human and machine, there are 2 that may seem very similar: collaboration and cooperation.

Co-funded by the European Union

In a collaboration, there is no a priori roles distribution, but a spontaneous roles distribution depending on the interaction history. In contrast, cooperation occurs when different roles are ascribed to the agents prior to the beginning of a task and this distribution is not questioned until its completion. While in collaboration the agents work on an even basis, cooperation has an uneven distribution of subtasks or roles during the task. Cooperating agents work towards the same end and need each other to complete the task but are not equal. In fact, cooperation is characterised by an asymmetric behaviour (Jarrassé et al., 2012).

In particular, collaboration can be seen under 3 different perspectives: the organisational perspective, the relationship perspective, and the interaction perspective. They correspond to different levels of deployment in human-machine collaboration, considering how humans and machines are organised, how they work together, and how they interact with each other. Specifically, the organisational perspective cares about forming the human-machine team organisations and solving task allocation problems. The relationship perspective analyses acceptance, trust, and dependence of human and machine on each other. Finally, the interaction perspective is mainly about designs of communication to foster mutual understanding and bilateral interventions via physical and mental interfaces (Xiong et al., 2022).

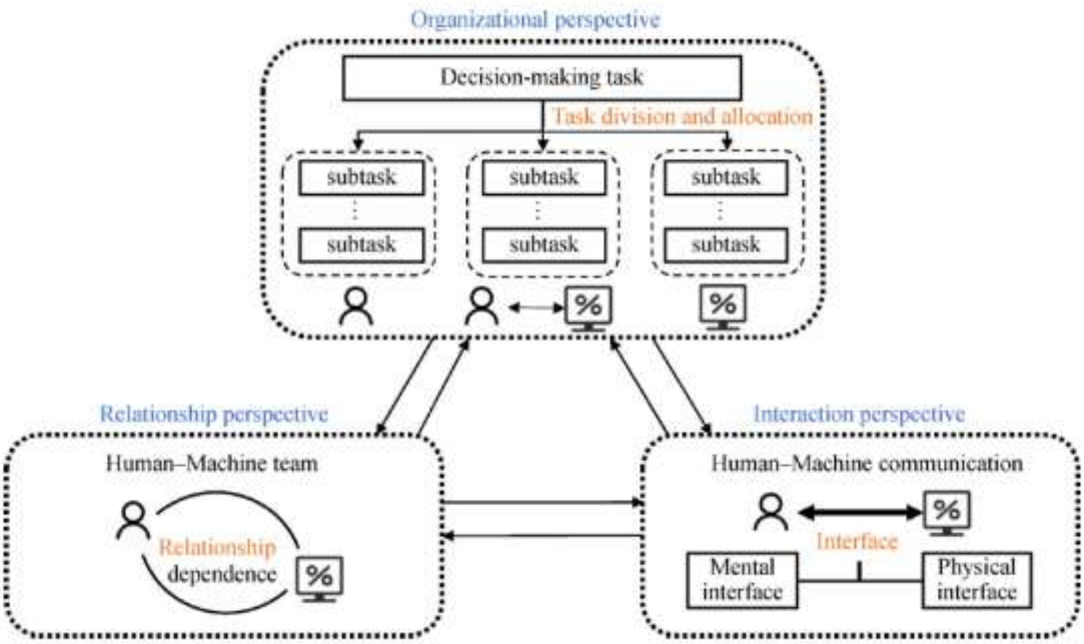


Figure 12 Perspectives on human-machine collaboration (Xiong et al., 2022)

A clear distinction is made in the "EASA Concept Paper: guidance for Level 1 & 2 machine learning applications":

- *Human-AI cooperation*: cooperation is a process in which the AI-based system works to help the end user accomplish his or her own goal.
  The AI-based system works according to a predefined task allocation pattern with informative feedback to the end-user on the decisions and/or actions implementation. The cooperation process follows a directive approach. Cooperation does not imply a shared situational

awareness between the end user and the AI-based system. Communication is not a paramount capability for cooperation.

- *Human-AI collaboration*: collaboration is a process in which the human end user and the AI-based system work together and jointly to achieve a common goal (or work individually on a defined goal) and solve a problem through a co-constructive approach. Collaboration implies the capability to share situational awareness and to readjust strategies and task allocation in real-time. Communication is paramount to share valuable information needed to achieve the goal, to share ideas and expectations.

In the EASA guidance, the distinction is useful to split the Level 2 of the AI applications (human/machine teaming) in 2 sub-levels:

- Human-AI cooperation: Level 2A AI.
- Human-AI collaboration: Level 2B AI.

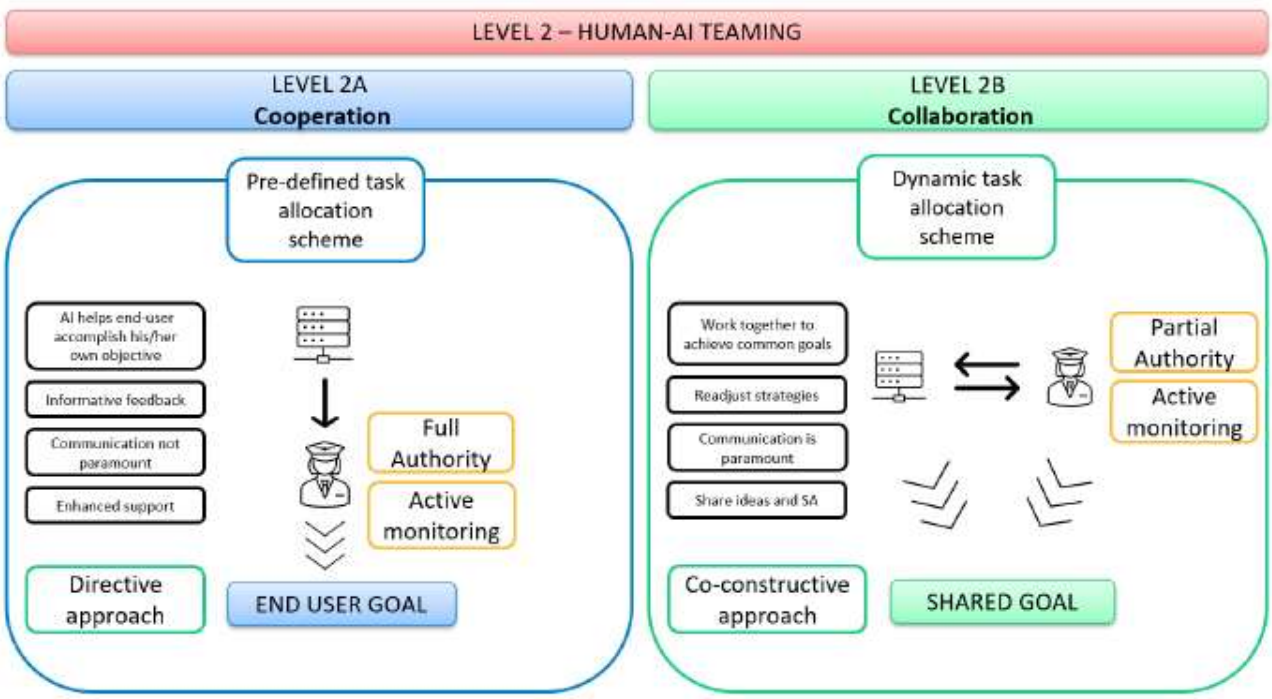


Figure 13 Human AI Teaming concept overview (EASA 2023)

4.3.1.2 The Theoretical framework in Risky Decision Making

Risky decision-making refers to the problem of making choices without knowing the exact consequences (Bier et al., 1999). In a typical risky scenario, the user has to deal with several choices, and each choice involves multiple possible outcomes. Thus, likelihoods and consequences are two critical dimensions to characterise the outcome of such a decision (Bedford & Cooke, 2001). In these kinds of scenarios, it is possible to notice cognitive biases in human decisions, and the use of simple heuristics to reach a solution (Kahneman & Tversky, 1979). Human-machine teaming for risky decision-making opens to many questions. For example, who should be assigned with which tasks, including cognition, judgement, and decision, and under what principles? How can a machine understand human decision-makers' values and behaviours and prescribe both normatively correct and subjectively acceptable solutions?

The opportunities for human-machine collaboration in risky decision-making can be characterised on the levels of uncertainty involved.

- On the one hand, when the decision task features low uncertainty, the research opportunities are mainly algorithm-centred, which lie in the effective utilisation of the computing power of machines (Patel et al., 2019).
- On the other hand, when the decision task is associated with higher uncertainty, the research opportunities become human-centred. High uncertainty makes many patterns in past data unaccountable, thus, the required complexity of algorithms increases to model and predict such data, and issues of overfitting and "black box" become vital (Topol, 2019; Amann et al., 2020).

Furthermore, in decision tasks with high uncertainty, the research opportunities lie in human–machine collaboration centred for two reasons. First, humans are vulnerable to various cognitive biases, and their capabilities of information processing are limited, whereas machines can calibrate biases and handle mass data in a consistent and normatively correct way. When human and machine judgments have disparity, machines should be able to explain why the human judgments are wrong. Second, machines are unable to handle highly uncertain and rare cases well. By contrast, humans can use intuition and experience to adapt to new situations and quickly learn and generalise reasoning across tasks.

In order to promote Human-Machine Collaboration in risky decision-making processes, the work of Xiong et al. (2022) proposes 3 challenges on how to organise Human-Machine Teams, enhance each other's capabilities, and facilitate mutual understanding and humans' trust in machines.

Challenge 1: Developing a more dynamic and flexible human-machine team organisation.
Designing the human-machine team organisation mode to make decisions under risk. Humans and machines undertake different roles in the environment and tasks with different levels of variability, uncertainty, and complexity (Daugherty & Wilson, 2018). The challenge can be divided into 3 parts.

- Applying a dynamic task allocation strategy to respond to dynamic characteristics and to support the combined performance. For specific risky decision-making tasks, human-machine teams may encounter multiple environment uncertainty risk levels and exhibit dynamic behaviours (Bier et al., 1999).
- Determining a fair distribution of the responsibilities in human-machine teams in risky decision-making. Risky decision-making always presents a number of negative outcomes.
- An appropriate accountability distribution in a human-machine team can affect acceptance and facilitate a beneficial human-machine relationship (Flemisch et al., 2012). Humans usually tend to blame the machine for the same mistake and negative outcomes (Dietvorst et al., 2015). This tendency would be more severe in risky decision-making with more uncertain negative outcomes.

To overcome Challenge 1, the following research questions must be considered.
(1) How should the human-machine team be organised and what are the criteria to decide which one (human, machine, or human-machine collaboration) holds the authority in risky decision-making?
(2) How should tasks between human and machine decision-makers, including cognition, judgement, and decision, be assigned? How can dynamic task allocation based on task requirements and the characteristics of human and machine decision-makers be achieved?

(3) What are the criteria to decide who should be accountable for the decision outcomes in human-machine teams in risky decision-making? How does a different accountability distribution impact the human-machine collaboration performance?

Challenge 2: Employing machines to help overcome humans' undesirable behaviours effectively (hence enhancing the human decision-maker) in risky decision-making.
Existing studies pay more attention to how machines assist humans for better decision-making than to leveraging machines to discover and correct human cognitive and behavioural limitations in risky decision-making. We break down the challenge into four parts.

- Determining the capability boundary of humans in risky decision-making. The capability boundary of humans is scoped by human cognitive and behavioural limitations in risky decision-making (Blumenthal-Barby et al, 2015).
- Developing adaptive machine design to support in overcoming or intervening humans' multiple limitations.
- In risky decision-making, behaviours of human decision makers, as well as multiple limitations in cognition and behaviour, are affected by multiple dynamic and uncertain factors (Cokely et al, 2009; Ordóñez et al, 2015).
- Evaluating the collaborative decision-making process objectively and subjectively. Evaluating the collaborative decision-making process can help understand the collaborative process and move the machine design and human-machine collaborative design forward (Damacharla et al, 2018).

To overcome Challenge 2, the following research questions are taken into account.
(4) What are human cognitive and behavioural limitations in risky decision-making? How can these limitations and their impacts be understood and modelled?
(5) How can machines provide normatively correct solutions for human cognitive and behavioural limitations? What impacts do different contexts or tasks have on human cognitive and behavioural limitations? In which way can machines be designed and developed to help overcome these limitations adaptively?
(6) What indicators can best describe and quantitatively evaluate the collaborative decision-making process?

Challenge 3: Developing communication and interface design to support mutual understanding and trust in human-machine teams.
Communication and information sharing have a critical role in achieving an understanding of intentions and behaviours and creating an effective human-machine team (Chen et al., 2018; Edmonds et al., 2019). More specific challenge details are described below.

- Design for intention identification and alignment. The identification, understanding, and alignment of respective goal(s), value(s), and intention(s) in a human–machine team can improve the efficiency and performance of Human-machine collaboration (Schaefer et al., 2017).
- Effective behaviour identification and monitoring of behavioural limitations in risky decision-making. When the capability boundary is known, monitoring and identifying the human's irrational behaviours or behaviours due to cognitive limitations are critical for the intervention toward the human (Damacharla et al., 2018).
- Appropriate intervention designs to overcome inconsistency in capabilities and behaviours in human-machine teams. When decisions of the human and machine decision-makers are

inconsistent or capability/behaviour limitations arise, appropriate intervention can effectively prevent possible negative outcomes (Daugherty and Wilson, 2018).
- Interaction design and evaluation considering human perception and understanding of machines. The physical interface has developed to be adaptive and algorithm dependent; more variables in the mental interface, such as trust and acceptance, should be considered to facilitate effective human–machine collaboration in risky decision-making (Dubois and Le Ny, 2020).

To overcome Challenge 3, we pose the following research questions for consideration.
(7) How do machines express their intentions, capabilities, and behaviours in risky decision-making? What behavioural indicators can represent human intentions? In which way can a human–machine team effectively align the goal, value, and intention?
(8) What behavioural indicators can represent the underlying cognition of human decision-making? How can machines identify and collect those indicators?
(9) How does a machine explain its decision-making rules? How does a machine understand humans' decision-making rules? How could the machine implement the intervention in an acceptable way?
(10) How can influencing factors in human–machine collaboration be modelled in risky decision-making? How can these models be embedded in algorithms behind the interaction interface?

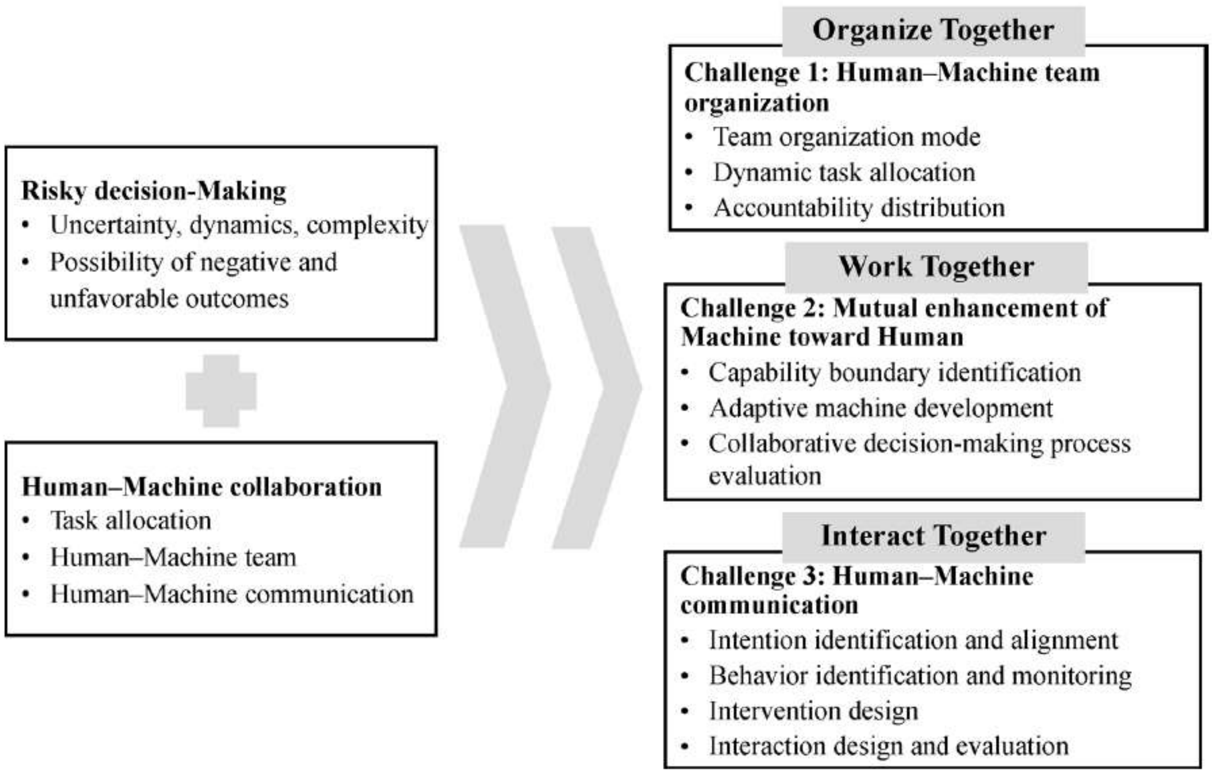


Figure 14 Challenges for Human-Machine Collaboration in risky decision-making (Xiong et al., 2022)

4.3.1.3 Task division and allocation

Effective human–machine (agent) interaction requires the appropriate allocation of indivisible tasks between humans and machines. Task allocation, recently also referred to as function allocation, decides which agent does what in a team. It represents an enabler for a successful interaction being a main topic of research in human-automation interaction (Ponti et al., 2022), especially when machines exhibit high levels of intelligence and autonomy. The topic is part of human-computer interaction and includes a large body of literature in different fields, including cognitive engineering and human factors. Function allocation covers human–human function allocation, team design, and human–automation function allocation. It is traditionally conducted as part of the human systems integration (HSI) process used during the design of complex systems (MILSTD-46855A, 2011) (Roth et al, 2019).

4.3.1.3.1 Classification of Task allocation methods

Consider different types of function allocation methods, such as fixed (static allocation) or dynamic (dynamic allocation). In static allocation, the functions allocated to the human and machine members are static and don't change based on situational factors. Dynamic allocation and reallocation of tasks between humans and machine agents involves integrating adaptive automation based on situational factors. It also includes the provision for human team members to seamlessly reassign tasks among themselves.

Multi-agents and human–agent task allocation methods can be classified into the following types:

- **homogeneous agent-based** is a task allocation method typically undertaken in structured environments, where all of the agents and tasks are of the same type and any agent can perform any task. Homogeneous task allocation is based on the assumption that all agents and performances across agents are identical, which is why these methods are usually applicable to multi-agent teams and not human–agent teams.
- **capabilities-based** is a task allocation method considering the heterogeneity of agents, commonly seeking to match the capabilities or types of agents with task demands. Heterogeneous agents vary in their capabilities, operating areas, and communication capabilities MABA-MABA (men-are-better-at, machines-are-better-at) is known as a classical theory outlining the general strengths of humans and machines and has been used as a basis for function allocation.
- **capacity-based** (or adaptive automation) is a task allocation method relying on human capacity information (e.g., workload, fatigue) to aid in the allocation of tasks (or level of automation control), aiming at keeping capacity in acceptable ranges.

4.3.1.3.2 Approaches for task allocation

A prominent approach used for years to decide which tasks are better performed by machines or by humans has been the HABA-MABA ("Humans are better at, Machines are better at") list firstly introduced by Fitts (1951). This list contains 11 "principles" recommending the functions that are better performed by machines and should be automated, while the other functions should be assigned to humans. Although researchers differ in what they consider appropriate criteria for task allocation, the influence of Fitts's principles persists today in the human factors' literature.

Co-funded by the European Union

Humans appear to surpass present-day machines in respect to the following:

1. Ability to detect a small amount of visual or acoustic energy
2. Ability to perceive patterns of light or sound
3. Ability to improvise and use flexible procedures
4. Ability to store very large amounts of information for long periods and to recall relevant facts at the appropriate time
5. Ability to reason inductively
6. Ability to exercise judgment

Present-day **machines** appear to surpass humans in respect to the following:

7. Ability to respond quickly to control signals and to apply great force smoothly and precisely
8. Ability to perform repetitive, routine tasks
9. Ability to store information briefly and then to erase it completely
10. Ability to reason deductively, including computational ability
11. Ability to handle highly complex operations, i.e. to do many different things at once

Figure 15 Fitts list (Fitts, 1951).

However, the HABA–MABA approach suffers from the clear limitation that the lists of what humans versus machines are better at can become quickly outdated as technologies continue to improve. In 2022 SEI (SEI, 2022) has identified a new list considering the AI-powered machine.

Humans surpass AI*

- **Exposing Bias**
- Identifying downstream impacts
- **Judgment**
- Recognizing Bias
- Responding to change
- Socio-political nuance
- Taking context into consideration

AI surpass humans*

- Computation
- **Computational complexity**
- Repetition
- Replication
- Scale
- Short and long-term memory
- Simultaneous operation
- Velocity

Figure 16 New Fitt's list (SEI, 2022)

Another limitation in the MABA approach is that by integrating the machine in performing a task, new tasks are created for the human who now has to interact with the technology (e.g., entering inputs, engaging/disengaging the automation, monitoring, etc.). These new tasks (e.g., monitoring system states and functioning) may, ironically, require what the Fitts report originally stated humans are bad at doing—namely, tasks requiring vigilance and little activity.

Another relevant approach for task allocation is represented by the Level of Automation (LoA) framework addressed in each sector. LoA frameworks incorporate taxonomies that specify which aspects of cognitive performance are being addressed (e.g., gathering the information, interpreting the information, generating solutions, deciding on action, taking an action), and the level of automation presumed. Much of the focus of LoA research is on understanding the impact of different LoA on human situation awareness, workload, complacency, trust, and ability to take over when automation fails.

(Roth et al., 2019) highlights that a major concern with the LoA approach as guidance to system designers is that tasks are described at too high a level of classification (e.g., information integration, decision, action). As a consequence, it limits the range of options for how work might be organised across the human and automated agents.

Cognitive Task Analysis (CTA) and Cognitive Work Analysis (CWA) methods are well suited for identifying and analysing the full range of demands of the work domain. CTA methods typically leverage knowledge of domain experts. CWA is an integrated set of analytic tools intended to represent the cognitive demands of work and the requirements to effectively support work performance. Work domain analyses are often conducted using an abstraction hierarchy (AH) representation of the goals, constraints, and functional means available to achieve the goals at different levels of abstraction.

Recent works have emphasised the importance of designing systems that enable more fluid distribution and redistribution of work to accommodate changing demands (Naikar, 2018); (Naikar & Elix, 2016); (Naikar, Elix, Dâgge, & Caldwell, 2017). Studies have shown that while team members may have formally defined roles and command structures, in practice, the allocation of tasks and leadership roles are more fluid, responding to the local demands of the situation. Accordingly, the idea is to analyse and design systems to support the functions that individuals and automated agents could, in principle, take on.

Interdependency analysis tools that represent both the human and the automated agent, the work to be performed, and the relationships between the human and the automated agent throughout the work have been provided. In analysing interdependencies, the tools consider not only new tasks emerging when automation is introduced (e.g., new monitoring tasks) but also ways that each agent can support the other agent in performing their tasks (e.g., a robot may need the help of a human to navigate around certain obstacles).

(Malone, 2018) proposes the concept of a Collective Intelligence further elaborating the need to allocate functions of cognitive processing, information flow, and task coordination beyond the scope or capability of individuals.

(Ali et al., 2022) proposes an interesting allocation method based on trust in both existing and novel tasks arriving at unknown times being different from task scheduling problems in which a set of tasks is known in advance such that they can be sequenced. Here trust is the willingness of the trustor to be vulnerable to the actions of the trustee.

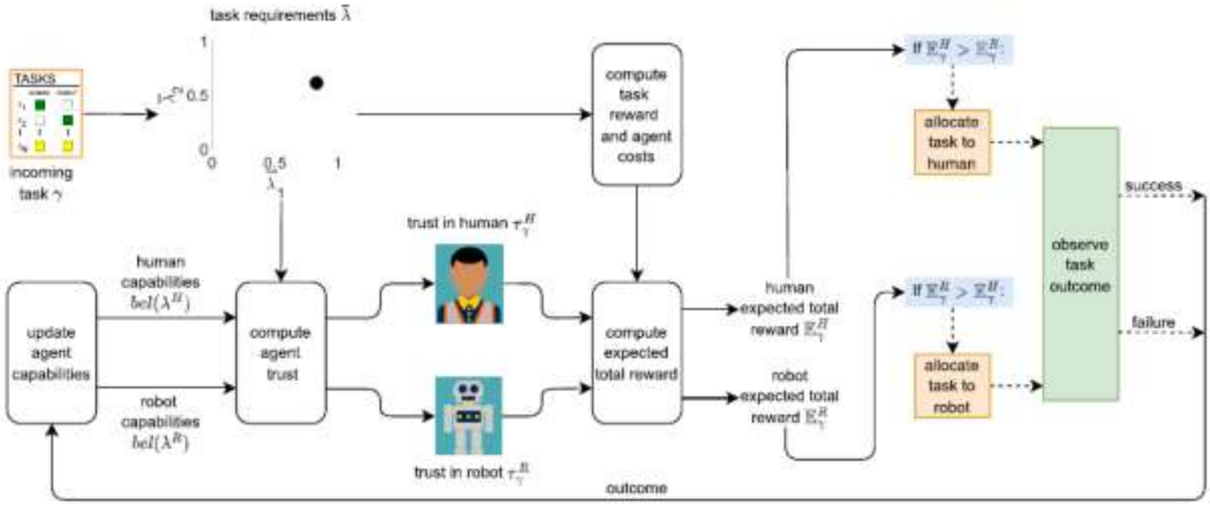


Figure 17 Flowchart with the main ideas of the artificial trust-based task allocation method for a team consisting of one human and one robotic agent. (Ali et al., 2022)

The process starts with an incoming task (black dot) defined by a set of task capability requirements. In this case, the incoming task is defined by two capability dimensions. The trust in each agent is computed using the capabilities belief distribution of that agent. The task reward and agent costs are computed using the task requirements. The expected total reward for each agent is computed using trust in the agent, task reward, and agent cost. The agent that maximises the expected total reward is allocated the task. The outcome of the task is observed as a success or a failure, which is used to update the capabilities belief distribution of the agent that executed the task. The process continues for each incoming task. In such method: 1) tasks are represented by the levels of capabilities required to successfully execute the task, and agents are represented by the levels of capabilities they possess; 2) trust in an agent to execute a new task can be reasoned about by considering the similarity between the new task capability requirements with existing tasks; 3) the belief in an agent's capabilities is updated over time as task outcomes are observed, either as successes or as failures; 4) the capabilities of an agent aren't known beforehand; 5) Task outcomes are not assumed to be strictly successes or failures, making them stochastic; 6) Task allocation is done using the robot's opinion (but the allocation can be done also by a third party).

Additionally, (Roth et al., 2019) highlights a set of factors that designers have to consider in designing a man-machine team:

- The need for coherence in the set of tasks that humans are assigned (i.e., avoiding "leftover" allocation).
- The need to avoid workload spikes as well as excessively low workload during long durations.
- The need to avoid situations where people are assigned responsibility for system outcomes, but the machine agent is assigned authority to automatically take action (i.e., avoiding authority/ responsibility mismatches).
- The need to avoid overly rigid (and unworkable) function allocations that lead to workarounds and disuse.
- The need to avoid brittle automation that is not reliable and/or fails abruptly when outside its boundary conditions.

- The need to avoid automation that results in excessive and untimely interruptions.

(Ali et al., 2022) addresses the following interesting dimensions of Task allocation: Novel tasks, Unknown and Dynamic Agent Capabilities, Negotiation and Fairness.

4.3.1.3.3 Novel tasks

Novel tasks that the human-robot team has not experienced before may occur, especially in dynamic situations. Part of the challenge in allocating novel tasks has to do with the difficulty in representing and characterising tasks. Very recent works address some approaches, but further research is needed since knowing the correct levels of capabilities to represent a task is a limitation of the current research.

4.3.1.3.4 Unknown and Dynamic Agent Capabilities

Agents on a human-robot team may be unfamiliar with the capabilities of their teammates if they have had limited interaction. A human can estimate the capabilities of another agent through interactions. Capabilities may grow through practice or training. Capabilities may also diminish if they are used infrequently or with fatigue. These concepts can have a great impact when dealing with AI-powered machines since on the one hand they stress the relevance of assuring training of the team to increase the acceptance and on the other hand they require a process of re-tuning of the machine after a potential degradation of performance.

4.3.1.3.5 Negotiation and Fairness

An agent may, for their own reasons, disagree with the agent responsible for task allocation. When such disagreements occur, agents will need a way to negotiate the allocation of a task until they reach a consensus. To start, the task allocation method will need to determine whether there are any disagreements among agents. One idea could be to simply request input when an agent disagrees with the allocation of a task. Once it is determined that disagreements between agents are present, how agents will negotiate and whether one agent will have the ultimate authority will have to be considered.

4.3.1.3.6 Metrics

(Sachendra Yadav, 2023) highlights that an important aspect of task allocation is to measure the effectiveness and efficiency of the adopted strategy. Identify the strengths and weaknesses of the strategy.

Measuring the effectiveness of a function allocation strategy in a human-AI team can be done through several metrics:

- **Workload:** Assess the workload of both human and AI agents to ensure it is balanced and manageable.
- **Stability of the Work Environment**: Evaluate how well the function allocation strategy adapts to changes in the work environment.
- **Mismatches Between Responsibility and Authority**: Identify any discrepancies between the responsibilities assigned to an agent and their authority.
- **Incoherency in Function Allocations**: Look for any inconsistencies or conflicts in the allocation of functions.
- **Interruptive Automation**: Measure the extent to which automation interrupts human work.

- Automation's Boundary Conditions: Determine the conditions under which automation performs effectively.
- **Function Allocations Limiting Human Adaptation to Context**: Assess whether the function allocation strategy restricts human adaptability to changes in context.
- **Workflow Performance**: Evaluate how well the function allocation strategy supports the overall performance of the workflow.

4.3.1.4 Elements of Trust

Trust is a subjective and abstract concept (Li et al., 2023), and is closely related to the fields of sociology and psychology. The definition of trust can vary significantly due to differences in the field of study, the specific objects and subjects being considered, and the contextual factors involved (Jøsang & McAnally, 2005). Therefore, there is no unanimous consensus on a single, widely accepted definition of trust. Typically, researchers define trust according to the specific scenario they are studying and identify the factors that influence it. Depending on the context, different forms of definition have been used. For example, in security (Internet of Things framework), trust can be understood as a relationship between nodes within a network. It can be defined as the subjective probability or possibility of one node exhibiting the desired behaviour as perceived by another node (Sfar et al., 2018). If the actions and behaviours of node B align with the expectations of node A, it can be said that node A trusts node B. In the context of node interactions, trust can be described as follows: Node B may be considered trustworthy by node A if node A believes that node B will strictly adhere to the expected and required behaviour.

Though there are many competing definitions of trust (Kaplan et al., 2023), there has not been a consensus on one specific definition of the concept (Sheridan, 2019). Trust has been examined through meta-analyses in relation to other forms of technology, such as automation (Schaefer et al., 2016) and robots (Hancock et al., 2011); (Hancock et al., 2021), and through systematic review (Hoff & Bashir, 2015).

There are different definitions not only in different research fields, but also in the same context. An overview of user trust definitions and influencing factors in human-AI interaction has been conducted (Bach et al., 2022). 23 articles have been analysed: 7 articles provided trust definitions; 8 articles conceptualised trust, but did not define it, and the remaining 8 articles neither defined nor conceptualised trust.

Concerning the articles that gave a definition of trust, 4 of them used Mayer's (Mayer et al., 2006) trust definition ((Foehr & Germelmann, 2020); (Glikson & Woolley, 2020); (Lin et al., 2019); (Thielsch et al., 2018)), 2 of them used Lee and See's (Lee & See, 2004) trust definition ((Hoffmann & Söllner, 2014); (Zhou et al., 2020)), 1 article developed its own definition in combination with different works (Yan et al., 2013).

The concept of trust is typically used in the technological environment. In the analysed articles, revised definitions are taken from the fields of sociology and psychology. The definitions are adapted depending on the objective of the research. Thus, there is the need to give an unambiguous definition that can be used in the context of the Human-Machine Interaction.

Title	Trust definition	Trust definition references
Alexa, can I trust you? Exploring consumer paths to trust in smart voice-interaction technologies (Foehr & Germelmann, 2020)	The willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party.	(Mayer et al., 2006)
Building e-commerce satisfaction and boosting sales: the role of social commerce trust and its antecedents (Lin et al., 2019)	The willingness of a party [the trustor] to be vulnerable to the actions of another party based on the expectation that the other [the trustee] will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party.	(Mayer et al., 2006)
Effects of personality traits on user trust in human–machine collaborations (Zhou et al., 2020)	The attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability.	(Lee & See, 2004)
Exploring trust of mobile applications based on user behaviors: an empirical study (Yan et al., 2013)	His/her belief on whether the application could fulfill a task as expected (the trustworthiness of mobile applications relates to their dependability, security, and usability).	Own definition and referenced (Avizienis et al., 2004)
Human trust in artificial intelligence: review of empirical research (Glikson & Woolley, 2020)	The willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party	(Mayer et al., 2006)
Incorporating behavioral trust theory into system development for ubiquitous applications (Hoffmann & Söllner, 2014)	The belief that an agent will help achieve an individual's goal in a situation characterized by uncertainty and vulnerability.	(Lee & See, 2004)
Trust and distrust in information systems at the workplace (Thielsch et al., 2018)	The willingness to depend on and be vulnerable to an Information System in uncertain and risky environments.	(Gefen et al., 2008; Mayer et al., 2006; Meeßen et al., 2020; Wang & Emurian, 2005)

Figure 18 Definition of Trust in various literature studies

Once the definition of trust has been established, there is a need to look for what elements can increase and/or decrease the human operator's trust in the AI-enabled system. A growing number of researchers argue that fostering and maintaining user trust is the key to calibrating the user-AI relationship, achieving trustworthy AI and further unlocking the potential of AI for society.

According to "EASA guidance for Level 1 & 2 machine learning applications", one of the main contributors in increasing the trust is explainability. As an example, if the explanation is warning the end user about the malfunction of the AI based system, the explanation will not positively influence the end user's trust in the system. Other influencing factors highlighted by the EASA guidance are:

- End user's general experience, belief, mindset, and prior exposure to the system.
- The maturity of the system.
- The end user's experience with the AI-based system, whether the experience is positive and there is a repetition of a positive outcome.
- The AI-based system knowledge on the end user's positive experience regarding a specific situation.
- The predictability of the AI-based system decision and whether the result expected is the correct one.
- The reinforcement of the reliability of the system through assurance processes.
- The fidelity and reliability of the interaction:
    - interaction will participate in end user's positive belief over the AI-based system's trustworthiness;
    - weak interaction capabilities, reliability, and experience can have a strong negative impact on the belief an end user may have in the trustworthiness of the whole system. It can even force him or her to turn off the system.

There are several factors that can influence the trust. They can be divided into 3 categories: factors related to the human user (the trustor), factors related to the AI-system (the trustee), and the contextual factors, related to the interaction between trustee and trustor, and the task to be performed. Furthermore, human-related factors can be divided into "users' abilities", such as situational awareness and task performance, and "personal characteristics", such as demographic information. AI-related factors can also be divided into "performance-related items", such as reliability, and "attributes", such as communication style. Contextual antecedents have here been related either to the "team", such as shared tenure, or "team tasking", such as difficulty. The analysis showed that human factors, AI factors and shared contextual factors are significant predictors of trust. Within each subcategory, multiple variables have both reported positive and negative influences on trust (Kaplan et al., 2023). Many of these have to do with the specific interaction between one single human user and a specific AI system.

Table 6 List of factors influencing trust in AI

Factors affecting trust					
Human (Trustor)		***AI (Trustee)***		***Contextual***	
User abilities	*Personal characteristic*	*Performance-related items*	*Attributes*	*Team*	*Team tasking*
Competency/ Understanding Expectancy Expertise Operator performance Prior experience Workload	Age Attitude towards AI Comfort with AI Culture Education Gender Personality traits Propensity to trust Satisfaction	Dependability Performance Predictability Reliability	AI personality Antropo-morphism Appearance Behaviour Communication Level of automation Reputation Transparency	Communication Interaction frequency Shared mental models Tenure	Risk Task complexity Task type

Monitoring factors that may influence trust in the AI system is undoubtedly important, but it remains a passive activity. It can be productive to identify which procedures can actively influence operator confidence. With this objective in mind, the study conducted by Bach et al. has taken into account 23 articles, from which were identified 3 main themes: socio-ethical considerations (8 articles), technical and design features (12), and user characteristics (22).

4.3.1.4.1 Socio-ethical considerations influencing user trust

An important task to enhance user confidence is the preparation and adaptation of the environment in which a system is to operate (Lee et al., 2021). This is because the development of AI-enabled

Co-funded by the European Union

systems is typically faster than the readiness of human users, and this mismatch might lead to low user trust. Some possible solutions to set up mechanisms in place to foster, maintain, and recover user trust (Binmad et al., 2017), might be, for example, ensuring user data protection, promoting high-quality user interactions and solution-oriented technical support. It was also suggested that user trust was likely to increase over time (Elkins & Derrick, 2013). Therefore, building and maintaining open communications with users, for example, by requesting ongoing feedback of an AI-enabled system being used, can be useful to increase user trust.

4.3.1.4.2 *Technical and* design *features influencing user trust*

During the development of a virtual agent whose task is to assist and communicate with a user, the following technical and/or design features were found to increase trust:

1. Anthropomorphism and human-like features, especially benevolent features (e.g., smiling, showing interest in the user) in an AI-enabled system.
2. Immediacy behaviours in which the AI-enabled system could create and project a perception of physical and psychological closeness to the user.
3. Social presence of the AI-enabled system (Morana et al., 2020); (Weitz et al., 2021).
4. Integrity of the AI-enabled system (i.e., repeatedly satisfactory task fulfilment) (Höddinghaus et al., 2021).
5. Supporting text/speech output when communicating with users.
6. Providing users with texts rather than a synthetic voice (Law et al., 2021).
7. A lower vocal pitch of the AI-enabled system.

Specifically, for AI/ML and automated algorithms, the following technical and/or design features were found to influence user trust:

1. Explanations and information regarding how the algorithm worked, AI's actions (Barda et al., 2020), (O'Sullivan et al., 2019), reflections of AI reliability, model performance (Zhang & Hußmann, 2021), feature influence methods, risk factors to predictive models, contextual information and interactive risk explanation tools (baseline risk and risk trends).
2. Correctness of AI/ML predictions.
3. AI/ML integrity.

In Thielsch et al., it was found that system reliability (dependability, lack and correctness of data, technical verification, distribution of the system) and the quality of the system information (credibility) influenced user trust. If an information system used a website to interact with users, multimedia features, security certificate/logo, contact information, and a social networking logo were found to be important for user trust (Sharma, 2015).

4.3.1.4.3 *User characteristics* influencing *user trust*

User characteristics can be divided in user inherent characteristics, user acquired characteristics, user attitudes and user external variables.

User inherent characteristics (personality traits and gender)

It was observed that personality traits of the users can influence predictive decision-making and trust in AI-enabled systems (Zhou et al., 2020). The study used the big five personality traits (Gosling et al., 2003) and found that Low Openness traits (practical, conventional, prefers routine) had the highest trust, followed by Low Conscientiousness (impulsive, careless, disorganised), Low Extraversion (quiet, reserved, withdrawn), and High Neuroticism (anxious, unhappy, prone to negative emotions). A user interface was suggested to include modules to identify and inform user personality traits to users. This would allow users to be aware of how their personality traits influenced their decision-making when interacting with an AI-enabled system. Moreover, women were more likely to trust an AI-enabled system (Morana et al., 2020).

User-acquired characteristics (user experiences and educational levels)

A previous experience with a provider or a producer of an AI-enabled system can influence user trust. Positive experiences with a system allowed the user to be rooted deeply in the provider's or producer's ecosystem, enabling the transfer of such trust to other systems from the same provider or producer. Generally, users without a college education were less likely to trust an AI-enabled system than those with a college education (Elkins & Derrick, 2013). The study also found that trust increased over time along with growing familiarity with the system, including when the initial trust level in the AI-enabled system is relatively low.

User attitudes (user acceptance and readiness, needs and expectations, judgement and perceptions)

User acceptance and readiness of an AI-enabled system were found to be key determinants of user trust ((Foehr & Germelmann & Germelmann, 2020); (Khosrowjerdi, 2016); (Klumpp & Zijm, 2019); (Smith, 2016)). Two studies suggested that addressing challenges such as artificial divide (Klumpp & Zijm, 2019) and user uncertainties (Hoffmann & Söllner, 2014) were fundamental for promoting user acceptance and readiness. The first study defined the artificial divide as the ability or lack thereof to cooperate successfully with AI-enabled systems. The study outlined that users might be divided by their motivation (e.g., intention to use) and technical competence toward AI-enabled systems. The study highlighted the importance of analysing artificial divide elements (e.g., rejection of an AI-enable system) and addressing challenges properly (e.g., early-stage user involvement, training, enhanced user experience and empowerment) to foster user trust and prevent mistrust.

The second study suggested that user uncertainties had to be addressed by identifying and prioritising the uncertainties and their antecedents in relation to a specific AI-enabled system, improving user understandability, sense of control, and information accuracy.

User needs and expectations of AI-enabled systems included user intention to use an AI-enabled system (Khosrowjerdi, 2016), relevance of technical system quality (e.g., reliability) and information quality (e.g., credibility) to users (Thielsch et al., 2018), as well as usefulness of an AI-enabled system to its users (Foehr & Germelmann, 2020). In general, user expectations of an AI-enabled system might not be aligned with the intention of the system's investors and developers (Lee et al., 2021). This might result in the system being operated in a way that was unforeseen by investors or developers, hitting and missing the target user expectations. The mismatch between user expectations and experiences was suggested to be a risk to user trust and needed to be addressed, especially when users were heavily dependent on specific AI-enabled systems.

For user judgement and perceptions, the key elements found to be affecting user trust in an AI-enabled system included perceived credibility (e.g., expertise, honesty, reputation, and predictability), risk (i.e., likelihood and severity of negative outcomes), and ease of use (e.g., searching, transacting and navigating) ((Corritore et al., 2012); (Foehr & Germelmann, 2020)) as well as perceived benevolence, integrity and transparency ((Elkins & Derrick, 2013); (Höddinghaus et al., 2021)). Importantly, it was found that the relatability a user felt to an AI-enabled system determined the user's trust in the system ((Thielsch et al., 2018); (Zhang & Hußmann, 2021)).

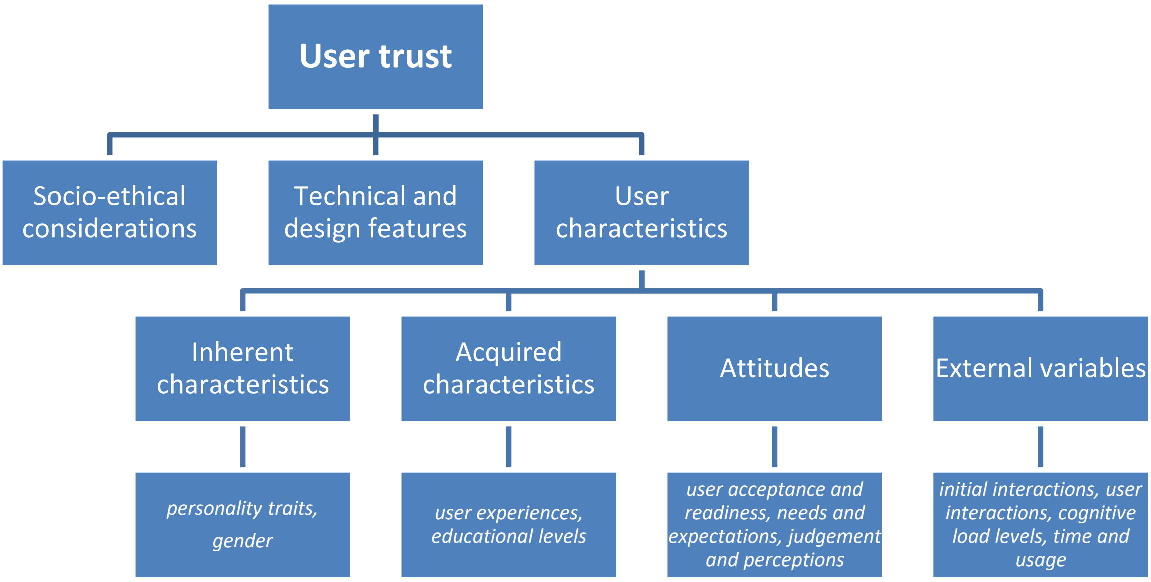


Figure 19 Influencing factors of trust

User external variables (initial interactions, user interactions, cognitive load levels, time and usage)

When an AI-enabled system was introduced to a potential user through the user's close relatives, friends or partner, the potential user typically used this opportunity to collect information regarding the system's benevolence, ability, and integrity (Foehr & Germelmann, 2020). Importantly, initial trust was likely to be fostered as well. In review-based recommender systems, the quality of user interactions on an AI-enabled system's platform was found to be a determinant of user trust ((Duffy, 2017); (Lin et al., 2019)). Creating an effective environment where users were willing to exchange social support and share high-quality reviews was suggested as crucial to foster and maintain user trust. Another important determinant of user trust was the user's cognitive load when interacting with an AI-enabled system (Zhou et al., 2020). When under a low cognitive load, the user was more willing to trust a system enabled by a greater availability of the user's cognitive resources which allowed more confidence and willingness to analyse and understand the AI-enabled system.

One study found that user trust increased as more time was spent interacting with an AI-enabled system (Elkins & Derrick, 2013), likely as a result of understanding the system better and thus perceiving it had greater integrity.

Finally, usage was suggested as a reliable predictor of user trust; the more a user used an AI-enabled system, the more they trusted the system.

Trust is one of the foundational attitudes within human interaction and without it, many important social bonds would be jeopardised. Without a minimal amount of trust in others, we would become paranoid and isolationist because of fear of deceit and harm (O'Neill, 2003).

The placement of trust in someone often requires a belief about their trustworthiness, but the two are not synonymous. Being trustworthy helps in gaining trust but is neither necessary nor sufficient. Deceivers can attract others' trust, so "Misplaced trust" is common enough. The trustworthy can be denied others' trust, so "Misplaced Mistrust" is also common enough.

It is possible to take into account three dominant trust paradigms to analyse if AI can be something that has the capacity to be trusted (Ryan, 2020):

- The rational account, in which the trustor is making a logical choice, weighing up the pros and cons, when determining whether to place their trust in the trustee; it is a rational calculation of whether the trustee is someone that will uphold the trust placed in them.
- The affective account, that states the trustor places a confidence in, and belief in, the goodwill of the trustee. There is an "expectation that the one trusted will be directly and favourably moved by the thought that someone is counting on him".
- The normative account, that implies the trustee's actions will be grounded on what he ought to do.

In order to evaluate whether the definition of trust is respected, it is possible to see if a set of characteristics are met:

1. 'A' has confidence in 'B' to do 'X'.
2. 'A' believes 'B' is competent to do 'X'.
3. 'A' is vulnerable to the actions of 'B'.
4. If 'B' does not do 'X' the 'A' may feel betrayed.
5. 'A' thinks that 'B' will do 'X', motivated by one of the following reasons:
    a. Their motivation does not matter (rational trust).
    b. 'B's' actions are based on a goodwill towards 'A' (affective trust).
    c. 'B' has a normative commitment to the relationship with 'A' (normative trust).

According to these characteristics, it can be noticed that AI can meet only the first three requirements. Furthermore, the first three requirements describe the rational account, thus the trust in AI can be seen more as a sort of reliance.

Concerning the fourth characteristic, the author highlights the difference between betrayal and disappointment. Betrayal closely relates to the confidence placed in, and confidence of, the trustee. Disappointment is the appropriate response when someone simply relied on someone or something to perform a task. We feel disappointed by those we rely on but feel betrayed by those we trust. The exclusion of betrayal is incompatible with the normative and affective accounts of trust, but non-necessarily the rational account of trust.

In conclusion, AI is not a thing to be trusted. The rational account of reliability does not require AI to have emotion towards the trustor (affective account) or be responsible for its actions (normative account).
One can rely on another based on dependable habits, but placing a trust in someone requires they act out of goodwill towards the trustor. This is the main reason why human-made objects, such as AI, can be reliable, but not trustworthy, according to the affective account.

In the normative account, the trustee must be held responsible for its actions, which AI cannot. Whereas, reliable AI places the burden of responsibility on those developing, deploying and using these technologies.

4.4 Explainable Artificial Intelligence (XAI)

The effectiveness of an AI-powered system is greatly limited by the machine’s inability to explain its decisions and actions to human users.

Automated decision-making systems always raise transparency and accountability issues. However, since the approval of the European General Data Protection Regulation (Reg. EU 2016/679) these problems have been addressed from a different perspective. The GDPR, indeed, rephrased and strengthened the individual prerogatives related to a “right to an explanation”, implicitly elevating the standards of compliance for systems involving opaque models and logics. Accordingly, researchers and companies started to develop new AI frameworks, putting more emphasis on the aspect of accountability related to these aspects

The need for XAI has to be seen considering 5 perspectives (Saeed, W. et al., 2023)

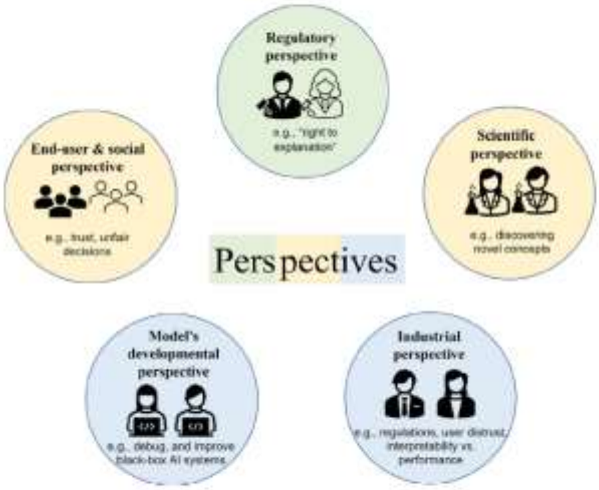


Figure 20 The five main perspectives for the need for XAI (Saeed, W. et al., 2023)

- **Regulatory perspective**: Black-box AI systems are being utilised in many areas of our daily lives, which could be resulting in unacceptable decisions, especially those that may lead to legal effects. Thus, it poses a new challenge for the legislation. The European Union’s General Data Protection Regulation (GDPR) is an example of why XAI is needed from a regulatory perspective. These regulations create what is called the ‘‘right to explanation”;
- **Scientific perspective**: XAI can be helpful to reveal the scientific knowledge extracted by the black-box AI models, which could lead to discovering novel concepts in various branches of science;
- **Industrial perspective:** Regulations and user distrust in black-box AI systems represent challenges to the industry in applying complex and accurate black-box AI systems. It can help in mitigating the common trade-off between model interpretability and performance, however, it can increase development and deployment costs;

- **Model's developer perspective**: XAI can be used to understand, debug, and improve the black-box AI system to enhance its robustness, increase safety, and minimise or prevent faulty behaviour;
- **End-user and social perspectives:** to understand if the system serves what it is designed for instead of what it was trained for. Figure 2 schematizes well the needs of accomplishing such perspective, due to its tight correlation with trust.

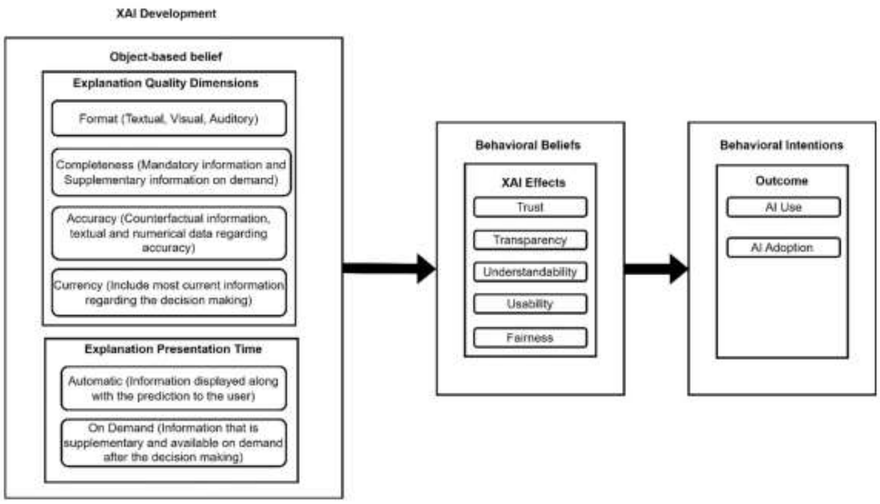


Figure 21 Synthesised framework for XAI research from a user perspective (Haque, A. B, et al., 2023)

In 2015, the Defence Advanced Research Project Agency (DARPA) launched a four-year research program on the topic with two main goals. The first one was to create machine learning techniques producing models that can be explained (their decision-making process as well as the output), while maintaining a high level of learning performance. The second goal was to convey a user-centric approach, in order to enable humans to understand their artificial counterparts.

Accordingly, the research topics around Explainability, first addressed by DARPA, then largely recognized by the scientific community ground on the following key questions (DW, 2019):

i) how to produce more explainable models
ii) how to design explanation interfaces
iii) how to understand the psychological requirements for effective explanations.

4.4.1 XAI Terminology

The first issue towards developing the ground knowledge of Explainability of Artificial Intelligence is the range of interchangeable terms used to describe some desired characteristics of an AI.

This includes (The Royal Society, 2019; Degas, A., et al., 2022):

- Interpretability, implying some sense of understanding how the technology works intending a property of an explanation;
- Transparency, implying some level of accessibility to the data or algorithm, indicating also the ability to be understandable to humans considering three kind of transparent models;

- "simulatable" models have the capacity to make humans understand their structure and functioning entirely;
- decomposable models can be decomposed into individual components, i.e., input, parameters and output, and their respective intuitions;
- algorithmically Transparent models behave "sensibly" in general with some degree of confidence.

- Justifiability, implying there is an understanding of the case in support of a particular outcome;
- Contestability, implying users have the information they need to argue against a decision or classification;
- Understandability, often termed as Intelligibility, implying a model that helps a user realise its functions. In other words, how the model works without any requirement of further explanation for the model's internal operations on the data;
- Comprehensibility, which has been used to define the ability of an ML model to represent its learned knowledge to humans in an understandable way. Clearly, the prior terms differ from the second on representing the internal operations on the data and the knowledge acquired from the data.

Above all, the term "Explainability" implies that a wider range of users can understand why or how a conclusion was reached.

Explainable AI (XAI) explains the inner process of a model i.e., used to provide the explanation of the methods, procedures and output of the processes and that should be understandable by the users.

DARPA adopted the term ''Explainable AI'' (XAI).

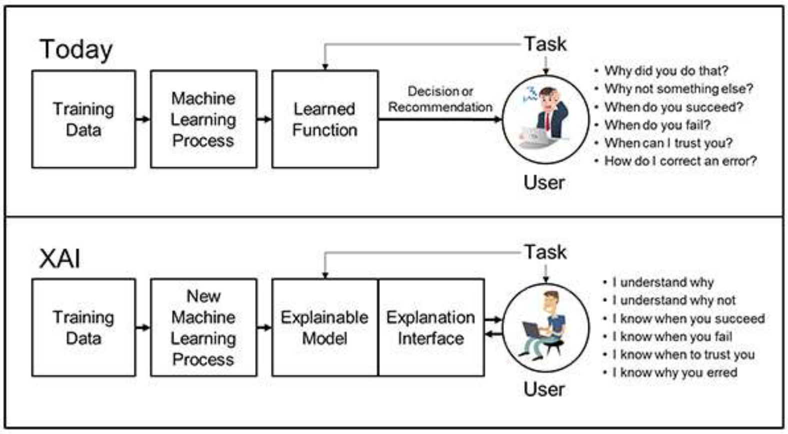


Figure 22 XAI Concept (DARPA-BAA-16-53, 2016)

According to (EASA, 2022) AI Explainability is defined as the

> "*Capability to provide the human with understandable, reliable, and relevant information with the appropriate level of details and with appropriate timing on how an AI/ML application produces its results".*

For EASA, different perspectives under which Explainability has to be studied lead to two different types of explainability: Development & post-ops explainability and Operational explainability.

4.4.2 XAI Principles and Attributes

The National Institute Standard Technology (NIST) in 2021 has introduced four principles (Phillips, P. et al., 2020) (Figure 1) to which an Explainable artificial intelligence (XAI) system has to adhere:

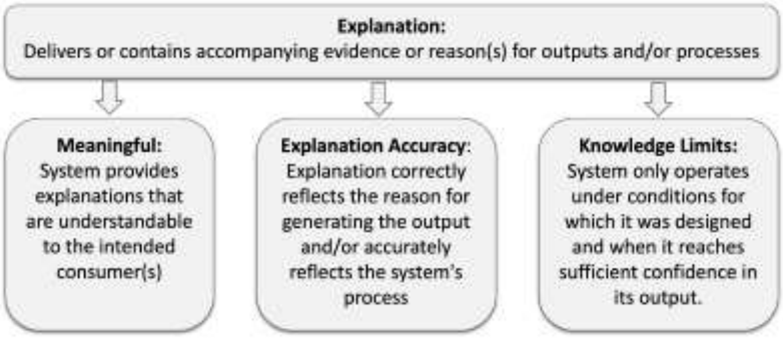


Figure 23 Illustration of the four principles of explainable artificial intelligence according to (Phillips, P. et al., 2020)

- **Explanation**: A system delivers or contains accompanying evidence or reason(s) for outputs and/or processes. By itself, the explanation principle is independent of whether the explanation is correct,
- informative, or intelligible. This principle does not impose any metric of quality on those explanations;
- **Meaningful**: A system provides explanations that are understandable to the intended consumer(s). This principle encompasses different perspectives – which are the intended target user of the explanation, what information people will find important, relevant, or useful, which is the needed prior knowledge and experiences to understand the explanation;
- **Explanation Accuracy**: An explanation correctly reflects the reason for generating the output and/or accurately reflects the system's process. Explanation accuracy is a distinct concept from decision accuracy. It mustn't be confused with Decision accuracy referring to whether the system's support to decision is correct or incorrect. The explanation may or may not accurately describe how the system came to its conclusion or action. Additionally, explanation accuracy needs to account for the level of detail in the explanation. For some target audiences and/or purposes, simple explanations could be sufficient for other could not.
- **Knowledge Limits**: A system only operates under conditions for which it was designed and when it reaches sufficient confidence in its output. The Knowledge Limits principle states that systems identify cases in which they were not designed or approved to operate, or in cases for which their answers are not reliable.

According to NIST Explanation are characterised by the following attributes:

- **Purpose** is the reason why a person requests an explanation or what question the explanation intends to answer;
- **Style** describes how an explanation is delivered:
  - level of detail as a range, from sparse to extensive
  - degree of interaction between the human and machine:

Co-funded by the European Union

- declarative explanations - the system provides an explanation, and there is no further interaction
- one-way interaction - explanation is determined based on a query or question input to the system
- two-way interaction - a conversation between people. The person can ask clarifying questions, or provide new avenues of exploration, and the machine answers

- Explanation format- visual and graphical, verbal, and auditory or visual alerts.

4.4.3 XAI Taxonomy

Various taxonomies are proposed for XAI. In (Speith, T., 2022 and Schwalbe, G., et al., 2023), based on extensive research, an XAI taxonomy is built considering different approaches:

- **The functioning-based approach**, meaning the way an explainability method extracts information from an ML model

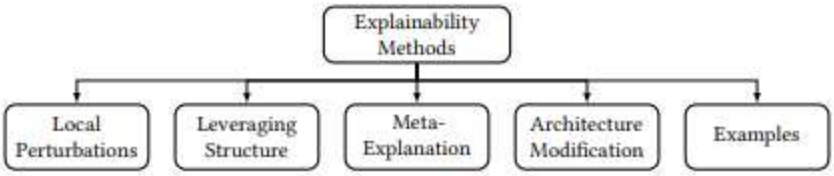


Figure 24 XAI taxonomy according to the Functioning based approach proposed by Samek and Muller (Speith, T., 2022)

- **The result-based approach** takes the result of an explainability method as the essential constituent for its classification

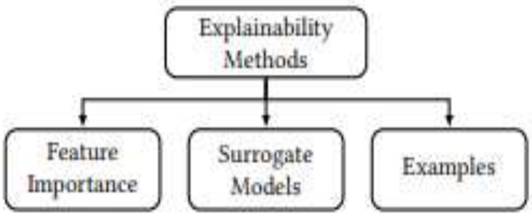


Figure 25 XAI taxonomy according to the Result based approach proposed by McDermid et al (Speith, T., 2022)

- **The conceptual-based approach** splits up the classification of explainability methods into several distinct conceptual dimensions such as: stage (ante-hoc vs. post-hoc), applicability (model-agnostic vs. model-specific), and scope (local vs. global). Such an approach has been integrated with other dimensions such as those linked to the output format distinguishing numerical, rules, textual, visual, and mixed. The stage represents the period in which a model generates the explanation for the output it provides. The stages are ante-hoc and post-hoc. Ante-hoc methods generally consider generating the explanation for the output from the very beginning of the training on the data while aiming to achieve the optimal performance. Post-hoc methods comprise an external or surrogate model and the base model. The base model remains unchanged, and the external model mimics the base model's behaviour to generate an explanation for the users. The post-hoc methods are again divided into two categories:

Model-agnostic and model-specific. The model-agnostic methods are applicable to any AI/ML model, whereas the model specific methods are confined to particular models.

- **The mixed based approach** joins the previous ones. An example is reported in Figure 7

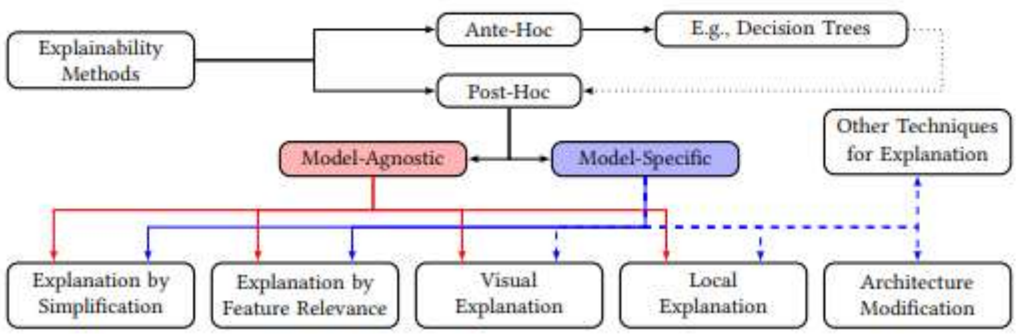


Figure 26 XAI taxonomy according to the mixed based approach (Speith, T., 2022)

After scanning more than 200 scientific articles published on XAI, Vilone and Longo deduced that the scope of explainability can be either global or local (Degas, A., et al., 2022).

In (Degas, A., et al., 2022) another classification of XAI is proposed, which seems to complement the functioning-based approach:

- **Descriptive XAI:** The system should be able to provide to all users the detailed description and rationale of the action to be taken.
- **Predictive XAI:** The XAI should be able to determine the 'what if' conditions or in other words, provide information to all stakeholders what will be the consequences of the actions that will be taken.
- **Prescriptive XAI:** The induced AI functions will, in addition to the above information, be able to suggest/propose the appropriate actions and options along with an appropriate explanation such that stakeholders can decide on the next course of actions. In the above scenario, the user can use the XAI prediction to assess the efficiency of potential actions—'what if'. XAI prescription will provide sufficient information to enable the user with immediate action to perform without testing them. One of the most recent work (Schwalbe, G., et al., 2023) presents a comprehensive taxonomy considering an increased number of dimensions including the task and other elements of the explanator (Figure 27)

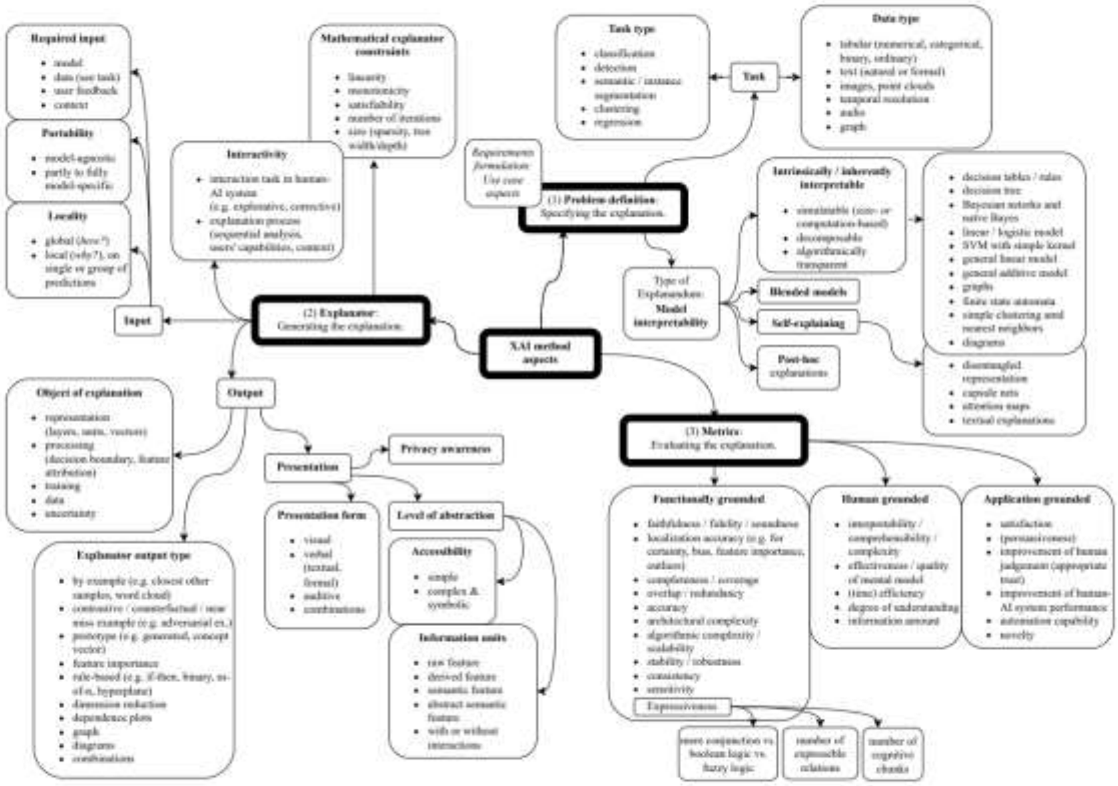


Figure 27 Overview of a complete taxonomy (Schwalbe, G., et al., 2023)

4.4.4 XAI Methods

The international community has developed a very broad range of different methods and approaches. Holzinger, A. et al., 2020, proposes an overview of the chronology of development of successive explanatory methods (see Figure 8) and a useful discussion on the basic ideas and the current limitations of the analysed methods.

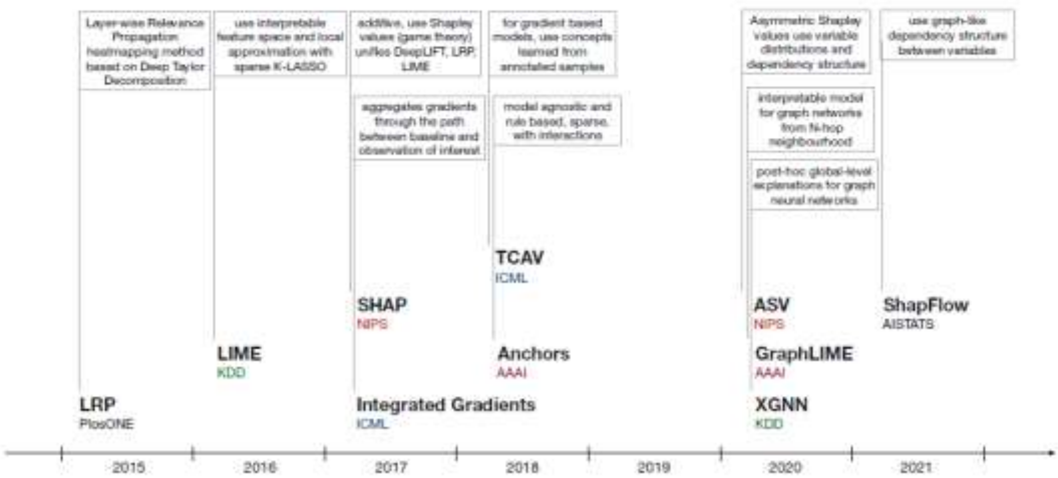


Figure 28 Chronology of the development of successive explanatory methods (Holzinger, A. et al., 2020)

In the table below are reported useful links for some of the most used XAI methods.

Table 7 Explainable AI Methods list and repositories

XAI method	GitHub Repo
LIME (Local Interpretable Model Agnostic Explanations)	https://github.com/marcotcr
Anchors	https://github.com/marcotcr/anchor
GraphLIME	https://github.com/WilliamCCHuang/GraphLIME
LRP (Layer-wise Relevance Propagation)	https://github.com/chr5tphr/zennit https://github.com/albermax/innvestigate
Deep Taylor Decomposition (DTD)	https://github.com/chr5tphr/zennit https://github.com/albermax/innvestigate
Prediction Difference Analysis (PDA)	https://github.com/lmzintgraf/DeepVis-PredDiff
TCAV (Testing with Concept Activation Vectors)	https://github.com/tensorflow/tcav
XGNN (Explainable Graph Neural Networks)	https://github.com/divelab/DIG/tree/dig/benchmarks/xgra ph/supp/XGNN
SHAP (Shapley Values)	https://github.com/slundberg/shap
Asymmetric Shapley Values (ASV)	https://github.com/nredell/shapFlex
Break-Down	https://github.com/ModelOriented/DALEX
Shapley Flow	https://github.com/nathanwang000/Shapley-Flow
Textual Explanations of Visual Models	https://github.com/LisaAnne/ECCV2016
Integrated Gradients	https://github.com/ankurtaly/Integrated-Gradients
Causal Models	No Github Repo
Meaningful Perturbations	https://github.com/ruthcfong/perturb explanations
EXplainable Neural-Symbolic Learning (X-NeSyL)	https://github.com/JulesSanchez/X-NeSyL, https://github.com/JulesSanchez/MonuMAI-AutomaticStyleClassification

(Degas, A., et al., 2022) proposed a rigorous review of the XAI methods, used in the last five years in ATM applications.

Methods for Explainability	Explanations				Stage		Scope		Design Space				Total Count
	N	*R*	*T*	*V*	*Ah*	*Ph*	*L*	*G*	*P*	*O/A*	*A*	*M/S*	
ANFIS	✓	✓	✓	✓		✓	✓	✓		✓	✓		9
Anchors	✓	✓	✓	✓		✓	✓	✓	✓				8
Feature Importance	✓		✓	✓		✓	✓	✓	✓				7
LIME	✓			✓		✓	✓	✓	✓		✓		7
RetainVis	✓		✓	✓		✓	✓	✓			✓		7
SHAP	✓			✓		✓	✓	✓	✓		✓		7
SRM				✓	✓		✓	✓		✓	✓	✓	7
SurvLIME-KS	✓			✓		✓	✓	✓	✓		✓		7
TreeExplainer	✓			✓		✓	✓	✓	✓			✓	7
BB-BC IT2FLS	✓	✓		✓	✓		✓					✓	6
CIE	✓		✓			✓	✓	✓	✓				6
ExNN				✓	✓		✓	✓	✓	✓			6
Methods for Explainability	Explanations				Stage		Scope		Design Space				Total Count
	N	*R*	*T*	*V*	*Ah*	*Ph*	*L*	*G*	*P*	*O/A*	*A*	*M/S*	
FDE	✓					✓	✓	✓		✓	✓		6
MAPLE	✓			✓		✓	✓	✓	✓				6
Generation			✓			✓	✓				✓	✓	5
GRACE			✓			✓	✓		✓		✓		5
HFS				✓	✓		✓	✓		✓			5
iNNvestigate				✓		✓	✓	✓	✓				5
J48		✓	✓		✓		✓	✓					5
Ada-WHIPS		✓				✓	✓		✓	✓			5
BN			✓	✓	✓		✓				✓		5
BRL		✓			✓			✓	✓		✓		5
CAM				✓		✓		✓	✓			✓	5
CFCMC			✓		✓			✓	✓		✓		5
CIT2FS			✓	✓	✓		✓			✓			5
Counterfactual Sets			✓			✓	✓		✓	✓			5
eUD3.5		✓			✓		✓	✓			✓		5
FINGRAM				✓	✓		✓		✓		✓		5
FormuCaseViz				✓		✓	✓				✓	✓	5
FURIA		✓			✓		✓		✓	✓			5
Ontological Perturbation		✓				✓	✓			✓	✓		5
RBIA				✓		✓	✓	✓		✓			5
RuleMatrix				✓		✓		✓			✓	✓	5
FFT		✓			✓		✓		✓				4
ICM			✓			✓	✓		✓				4
LORE		✓				✓	✓		✓				4
MTDT				✓	✓			✓	✓				4
Mutual Importance	✓					✓	✓		✓				4
OC-Tree		✓			✓			✓	✓				4
Attention Maps				✓		✓	✓		✓				4
Causal Importance	✓					✓		✓			✓		4
CTree		✓			✓		✓		✓				4
TCBR			✓			✓	✓				✓		4
Template-based Natural Language Generation			✓		✓		✓		✓				4
TREPAN		✓				✓	✓				✓		4
WM Algorithm		✓			✓			✓				✓	4
al Count	14	15	15	23	18	28	38	26	27	11	20	8	–

Figure 29 XAI methods with associated types of explanations (N: Numeric, R: Rules, T: Textual, V: Visual), stage (Ah: Ante-hoc, Ph: Post-hoc), scope (L: Local, G: Global) of explainability, and the design spaces (P: Prediction, O/A: Optimisation/Automation, A: Analysis, M/S: Modelling/Simulation) (Degas, A., et al., 2022)

(Schwalbe, G., et al., 2023) proposed another rigorous review of the XAI methods independent of the application domain that goes through the three strategies proposed by the DARPA program: deep explanation, interpretable models and model induction.

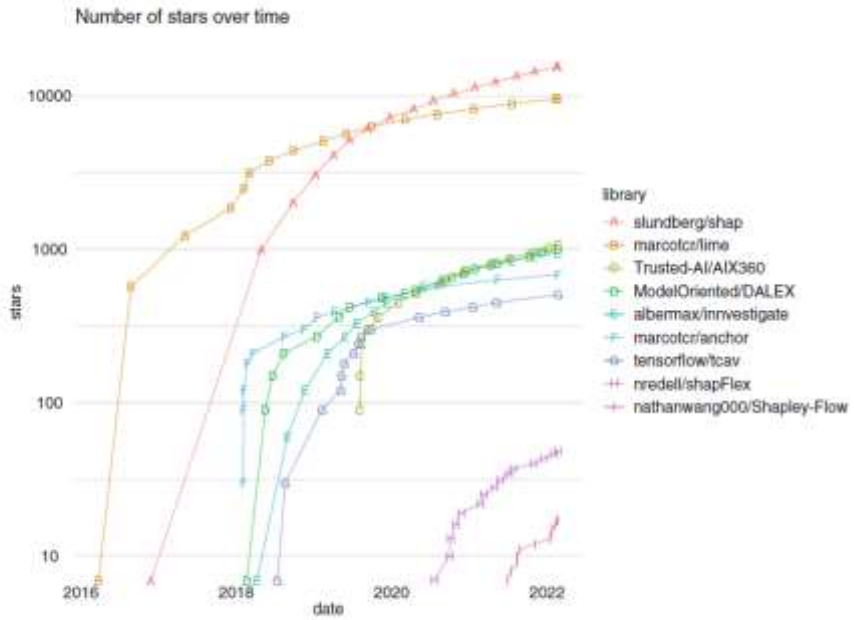
Further analysis on the methods is provided in (Holzinger, A. et al., 2020), a complete book of a conference where the term XXAI means beyond the Explainable AI.

An interesting summary of limitations of the current "explainers" is reported in (Swamy, V. et al., 2023). Post-hoc approaches are most commonly investigated, and they don't impact the model accuracy and don't require additional effort during training. Local, instance-specific post-hoc techniques such as LIME and SHAP have been effectively utilised in a variety of models. Counterfactual explanations have been used in numerous classification tasks. Each of the post-hoc XAI solutions presented above, among many others not mentioned, have weaknesses for deployment in a real-world setting. The computational time, especially with SHAP, LIME, or counterfactual generation, is in the tens of minutes; not real time enough for users. In most cases, there is no measurement of trust or confidence in a generated explanation. The actionability and human-understandability of the explanation is based on the input format. As human-centric tasks often use tabular or time series data, their subsequent explanations are often not concise, actionable or interpretable. Lastly, the consistency of the explanations is not intrinsically measured; generating an explanation for the next step in the time series could vary greatly from the previous step. Several explainability methods could produce vastly different explanations with different random seeds.

Less research has focused on in-hoc methods.

4.4.5 XAI main Toolboxes

The international community has developed a very broad range of different methods and approaches and here we provide a short concise overview to help engineers but also students to select the best possible method. Figure 10 shows some of the most popular XAI toolboxes (Holzinger, A. et al., 2020).



EUROPEAN PARTNERSHIP
Co-funded by the European Union

Figure 30 Number of stars on GitHub for the most popular repositories (Holzinger, A. et al., 2020)

While these repositories focus on the explanation task, the new Quantus toolbox proposed in (Hedström, A., et al., 2022) offers a collection of methods for evaluating and comparing explanations.

XAI for reinforcement learning tasks There are some studies specifically focusing on explanations in tasks solved by reinforcement learning. One is that by (Puiutta E, 2020.) It reviews more than 16 methods specific to reinforcement learning in a beginner-friendly way. Comparable and more recent surveys on the topic are by Heuillet et al (2021) and (Vouros, 2022).

XAI from a HCI perspective When humans interact with AI-driven machines, this human-machine-system can benefit from explanations obtained by XAI. DARPA program structure anticipated the need for a grounded psychological understanding of explanation, summarising psychological theories of explanation to assist the XAI developers and the evaluation. The concept of user-centric XAI requires a highly interdisciplinary perspective. This is based on fields such as computer science, social sciences as well as psychology in order to produce more explainable models, suitable explanation interfaces, and to communicate explanations effectively under consideration of psychological aspects. Figure 11 illustrates a top-level descriptive model of the XAI explanation process (Gunning, D et al., 2021)

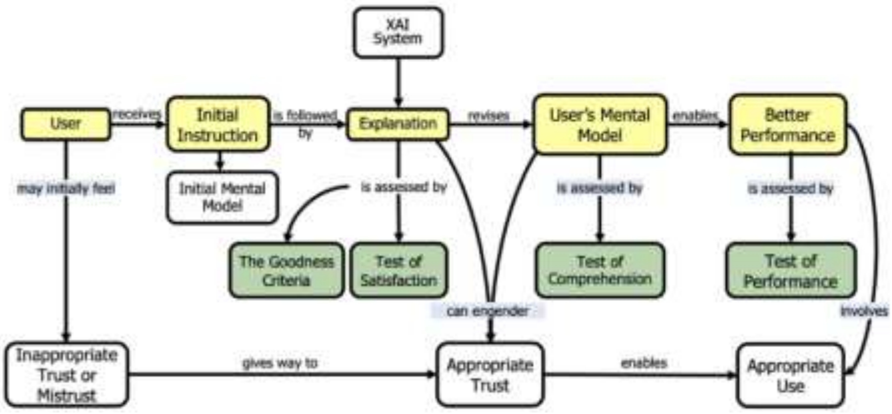


Figure 31 Psychological model of explanation. Yellow boxes illustrate the underlying process, the green boxes illustrate the measurement opportunities and the white boxes illustrate potential outcomes (Gunning, D et al., 2021)

The work initiated by DARPA has been further developed. Hence, by now there are several surveys concentrating on XAI against the background of human-computer interaction (HCI). The survey in (Ferreira JJ, et al., 2020) is slightly longer, and may serve as an entry point to the topic for researchers. (Schwalbe, G., et al., 2023) highlights different works exploring the concept of explanations and explainability concluding that: (a) (local) explanations should be understood contrastively, i.e., they should clarify why an action was taken instead of another; (b) explanations are selected in a biased manner, i.e., do not represent the complete causal chain but few selected causes; (c) causal links are more helpful to humans than probabilities and statistics; and (d) explanations are social in the sense that the background of the explanation receiver matters. Other works rigorously develop a taxonomy for evaluating black-box XAI methods with the help of human subjects with concrete suggestions for study design. Very recent work on XAI metrics is provided by (Mueller ST, et al., 2021.) where concrete and practical design principles for XAI in human-machine-system and several relevant XAI metrics are recapitulated.

DARPA studies on such aspects (Gunning et al., 2021) were blocked by the recent pandemic, but the main conclusions on such aspects were:

- Users prefer systems that provide decisions with explanations over systems that provide only decisions. Tasks where explanations provide the most value are those where a user needs to understand the inner workings of how an AI system makes decisions.
- To improve user task performance, the task must be difficult enough that the AI explanation helps (PARC, UT Dallas).
- User cognitive load to interpret explanations can hinder user performance. Combined with the previous point, explanations and task difficulty need to be calibrated in order to improve user performance.
- Explanations are more helpful when an AI is incorrect and are particularly valuable for edge cases
- Measures of explanation effectiveness can change over time.

XAI from the evaluation perspective. Another hot research topic around XAI are metrics for measuring the quality of explanations for human receivers. In 2017, Doshi-Velez and Kim proposed a base work on XAI metric categorization (Doshi-Velez et al., 2017). DARPA proposed metrics as well (DW, 2019)

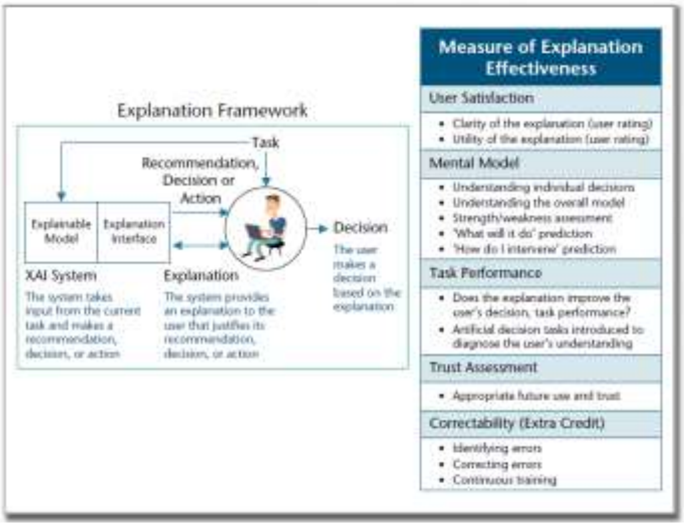


Figure 32 Measure of Explanation proposed by DARPA (DW, 2019)

(Schwalbe et al., 2023) collects from other surveys latent dimensions of interpretability with recommendations and proposes a classification metrics into:

- **Functionally grounded**: Faithfulness fidelity soundness or causality, Completeness, or coverage, Localization accuracy Overlap, Accuracy of the surrogate model, Architectural complexity, Algorithmic complexity, Stability or robustness, Consistency, Sensitivity, Expressiveness or the level of detail;
- **Human grounded:** Interpretability or comprehensibility, Effectiveness, (Time) Efficiency, Degree of understanding, Information amount;
- **Application grounded:** Satisfaction, Persuasiveness, Improvement of human judgement, Improvement of human-AI system performance, Automation capability, Novelty.

A full focus on metrics for evaluating XAI methods is set in the recent work by (Zhou J, et al., 2021).

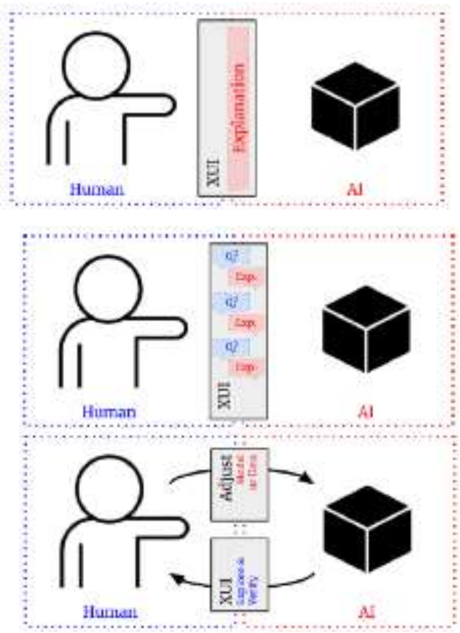
4.4.6 XAI from Explanation Interface perspective (Chromik et al., 2021)

The DARPA program distinguishes between the explainable model and the explanation user interface disentangling the ML model behaviour analysis from communicating it to the user. DARPA classified the interfaces according to the terms reported in the figure below.

Explanation Interface
Reflexive and Rational
Narrative Generation
3-Level Explanation
Acceptance Testing
Interactive Training
XRL Interaction
Show-and-Tell Explanations
Argumentation and Pedagogy
Decision Diagrams
Interactive Visualization
Bayesian Teaching

Figure 33 Explanation Interface types (Gunning, D et al., 2021)

(Chromik, M., et al., 2021) defines an XAI Explanation interface with the term XAI user interface (XUI) as the sum of outputs of an XAI system that the user can directly interact with. Most XAI research focuses on computational aspects of generating explanations while limited research is reported concerning the human-centred design of the XUI. Google's People+AI Guidebook2 presents case studies of explanations integrated into mobile apps. The relevant point highlighted in literature is that *"…explainability can only happen through interaction between human and machine..". In (Chromik, M., et al., 2021)* is argued that different types of interaction of AI with Human lead to different type of explanation interface proposed that on turn identify further research subtopics:

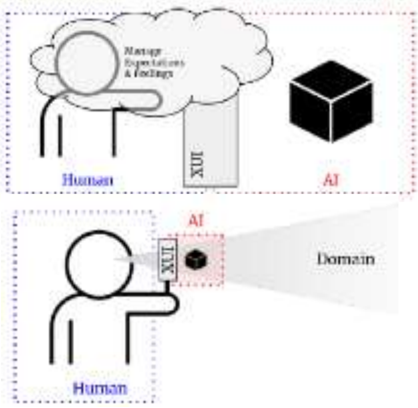


- **Interaction as (Information) Transmission -** this interaction is mainly about unidirectional XUI presents users with accurate or complete explanation about AI behaviour;
- **Interaction as Dialogue –** XUI allows a cycle of communication of inputs/outputs by the computer and perception/action by a human;
- **Interaction as Control** Built on control theory XUI feeds control signals from the ML model to the human controller (feedback). These inform the controller how to change parameters of the ML model or its data so that the model adjusts its behaviour (feedforward);
- **Interaction as Optimal Behaviour:** XUI provides explanations for training humans to have better interactions with AI, for example, when they face erroneous AI systems or exhibit misconceptions caused by cognitive biases. It is divided in research that (i) examines limitations that occur during the interaction with an XAI and (ii) designs interactions to better moderate these limitations;

EUROPEAN PARTNERSHIP

Co-funded by the European Union

- **Interaction as Experience -** XUI emphasises managing the expectations and preferences of users about the AI. It centres around the explanatory goals of trust;
- **Interaction as Embodied Action** XUI is built on theories from the computer-supported cooperative work (CSCW) community, such as mutual goal understanding, pre-emptive task co-management and shared progress tracking. XUIs which are not only about understanding AI agents (interaction as transmission), but which enabled them to also influence the agents' actions;
- **Interaction as Tool Use -** XUI is built on Activity theory where the system influences the mental functioning of individuals and facilitates the learning from AI.



4.5 Performance Measures

In addition to the key performance indicator (KPI) defined SESAR Performance Framework other metrics have been proposed in the literature. As long as we do not completely replace humans with unmanned autonomous systems the best choice is human machine teaming or collaboration, but such teaming comes with its own set of challenges. Metrics are crucial to accurately and effectively measure human-machine teaming (HMT) across multiple fields. As stated in (Damacharla et al., 2018), to be deemed as an HMT, a team should contain at least one human and one machine/intelligent system and they propose to define HMT as a combination of cognitive, computer, and data sciences; embedded systems; phenomenology; psychology; robotics; sociology and social psychology; speech-language pathology; and visualisation, aimed at maximising team performance in critical missions where a human and machine are sharing a common set of goals. The machines that take part in an HMT must belong to one of the following categories: unmanned aerial vehicles (UAVs), unmanned ground vehicles, AI robots, digital assistants, and cloud assistants. Following this approach, metrics can be related to human, machine and team.

4.5.1 Human metrics

Human-metrics measure different human aspects such as system knowledge, performance, and efficiency that can be used to evaluate a human agent in an HMT. *Situational awareness* (SA) is measured by monitoring task progress and sensitivity to task dynamics during execution. The degree of *mental computation* estimates the amount of *cognitive workload* an operator manages to complete a task (Steinfeld et al., 2006), (Sammer et al., 2007). The accuracy of a mental model of an operator depends on interface comprehensiveness and simplicity in addition to control and compatibility a machine provides. *Attention allocation* measures the attention an operator pays to a team's mission and the operator's ability to assign strategies and priorities of tasks dynamically. The metric also considers an operator's degree of attention over multiple agents. It is measured using eye tracking, duration of eye fixations to an area of interest, and task completion rate, while *attention allocation efficiency* is measured using wait times (Pina et al., 2008), (Crandall et al., 2007). *Intervention frequency* is the frequency with which an operator interacts with the machine (Harriott et al., 2014). As per the literature, operators' intervention frequency is also known as *intervention rate* or *percentage requests*. *Stress* can be physical or mental. However, both may indicate the operator's workload and are measured in two ways. First, researchers perform sample testing of humans' stress hormones, such as cortisol and catecholamine, which are found in blood, saliva, and urine samples (Rapolienė et al., 2016). Second, researchers can perform a detrended fluctuation analysis of a human's heartbeat (Yazawa et al., 2013).

Human *safety metrics* involve evaluation of the risk posed to the human life while working near machines, for example, the location of the machine relative to the human. Human factor studies suggest that humans can establish the best cooperation with a machine through a 3D immersive environment (Corbillon et al., 2017). In (Forouzandehmehr et al., 2013) and (Saad et al., 2016), researchers suggest that humans can be more effective when the environment and goals are in their best interest. Other human performance attributes such as psychomotor processing, spatial processing, composure and perseverance are important to improve the team cohesion through human performance enhancement.

4.5.2 Machine metrics

Machine-level metrics related to HMT cope with efficiency, performance and accuracy. In particular, machine *self-awareness*, or the degree to which a machine is aware of itself (limitations, capacities), is a precursor to reducing the human cognitive load and measured based on *autonomous operation time*, *the degree of autonomy* and *task success* (Steinfeld et al., 2006), (Gorbenko et al., 2012). Instead, *unscheduled manual* operation time may either be an interruption period in the current plan execution or an unexpected assigned task (Schreckenghost et al., 2010). *Neglect tolerance* (NT) is interpreted in various ways, such as machine performance falling below expectation, time to catch-up, the idle period, or operation time without user intervention. *State metric* helps track the machine or plan state based on four dynamic states: assigned, executed, idle, and out of the plan. *Robot attention demand* (RAD) is a measure of the fractional ''task time'' a human spends to interact with a machine. *Fan out* (FO) is a measure of how many robots with similar capabilities a user can interact with simultaneously and efficiently and is the inverse of RAD (Abou Saleh, 2010). *Interaction effort* (IE) is a measure of the time required to interact with the robot based on experimental values of NT and FO and is used to calculate RAD (Abou Saleh et al., 2010), (Crandall et al., 2005).

4.5.3 Team metrics

Mission assignment and execution is the key focus of team metrics. Task difficulty represents the mental load a particular task generates (Greitzer et al., 2005). The *task difficulty* metric for a machine depends on FO and requires three factors for measurement: *recognition accuracy*, *situation coverage*, and *critical time ratio* of a machine (Glas et al., 2011). *Recognition accuracy* is the ability of the machine to sense its I/O parameters. *Situation coverage* (SC) is the percentage of situations encountered and accomplished by the robot. SC is defined based on plan and act stages of the mission. *Critical time ratio* is the ratio of time spent by a robot in a critical situation to the total time of interaction (Glas et al., 2011). *Network efficiency* is the rate of flow of information between the human and the machine and determines the efficiency of interaction. It also influences time taken for scheduled and unscheduled manual operations, accuracy of mental computation, negligence tolerance, and human-machine ratio (Harriott et al., 2014). Four well-known subclasses of *false alarms* are true positive (TP), true negative (TN), false positive (FP), and false negative (FN) (Elara et al., 2010). While false alarms measure complex communication between humans and machines in a team, people may ignore false alarms. A human factor study presented a trade-off between ignoring false alarms and *misses* and concluded that alarms are strongly situation dependent (Meyer et al., 2001). Some other team metrics that can be used in effective interactions are *hits*, *misses*, *automation bias* and *misuse of automation* or metrics based on application scenario (Doisy et al., 2014). *Robustness* measures the ability of the team to adapt to the changes in task and environment during task execution (Shah et al., 2007) while *productivity* measures productive time compared to total invested time. *Task success ratio* indicates the number of completed versus allocated tasks (Schreckenghost et al., 2010).

4.5.4 Metrics classification

According to the review (Damacharla et al., 2018), metrics can be functionally classified in four classes: efficiency, time, mission and safety. Metrics to evaluate efficiency will give the observer the required V&V to tune each agent to operate with maximum efficiency (e.g. *attention allocation*, *decisions accuracy*, *mental workload*). Time metrics provide data related to the time taken for different operations by machine, human, and team, and these metrics are very important in decision-making and performance and status determination (e.g. *neglect tolerance*, *critical time ratio*, *autonomous operation time*). Mission metrics measure attributes related to a task such as planning (e.g. *reliability*, *trust, total coverage*). Safety metrics measure the agent and mission safety during task execution (e.g. *risk to human*, *general health*, *critical hazard*). Another class of metrics, termed as applied metrics, deals with the practicality and research on metrics and is divided into research (e.g. *fatigue*, *stress*, *situation awareness*) and non-research (e.g. *false alarms*, *team productivity*, *task success*) metrics.

Subjective metrics (SM) are used to measure abstract qualities based on human perception. These metrics may include feedback or judgement from observers (superiors or experienced professionals), for example, self-feedback, evaluation, or ratings. These metrics are measured using scales rating from experts.

Objective metrics (OM) are task-specific tools, functions, and formulae to measure task performance quantitatively. OM are developed to measure an activity that can be changed, customised, or expressed by a value for comparison. Finally, some metrics can be measured in *real-time*, others can be evaluated only after the accomplishment of the mission.

4.6 Current Development in Aviation

4.6.1 Projects in the last five years

The following table represents the summary of projects that have dealt with Artificial Intelligence and with Digital Assistants in the last five years. Most of them are funded under the SESAR Program, fewer under Horizon Europe. Very few projects were recovered that were developed in extra-European Countries. Sic projects address Digital Assistant and Human-Ai teaming aspects.

		Project name	Starting date	Ending date	Coordinating entities	Summary
SESAR JU	**First projects**	**MALORCA**	01-04-2016	31-03-2018	Deutsches Zentrum für Luft-	The project proposed a general, cheap and effective solution to automate the speech recognition, adaptation and

					und Raumfahrt	customization process to new environments, taking advantage of the large amount of speech data available in the ATM world.
		INTUIT	01-03-2016	28-02-2018	Nommon Solutions and Technologies	The purpose of the project was to explore the potential of visual analytics, machine learning and systems modelling techniques to improve the understanding of the trade-offs between ATM Key Performance Area (KPAs), identify cause-effect relationships between Key Performance Indicators (KPIs) at different scales, and develop new decision support tools for ATM performance monitoring and management.
		DART	17-06-2016	19-06-2018	Ecole Nationale de l'Aviation Civile (ENAC)	The project aimed to understand the suitability of applying big data techniques for predicting multiple correlated aircraft trajectories based on data driven models and accounting for ATM network complexity effects.
		ARTIMATION	01-01-2021	31-12-2022	Malardalens Universitet	The project has introduced innovative AI methods to predict air transportation traffic and to optimise traffic flows based on the domain of explainable artificial intelligence. ARTIMATION aimed to ensure safe and dependable decision support, focusing on transparent AI models that include visualisation, explanation and

						generalisation with adaptability over time.
	Other projects	**TAPAS**	01-06-2020	30-11-2022	Centro de Referencia Investigacion Desarrollo e Innovacion ATM (CRIDA)	The main objective was to provide a set of principles and criteria which pave the way for the deployment of AI/ML-based technologies in ATM in a safe and trustworthy manner. eXplainable Artificial Intelligence (XAI) techniques, together with Visual Analytics, can help to explore trade-offs between efficiency of AI implementations and the suitability for deployment in specific applications.
		MAHALO	01-06-2020	30-11-2022	Deep Blue	The project aimed to design an automated AI, ML and deep neuronal learning-based explainable system for problem solving between aircrews and air traffic controllers. Trained by the individual operator, the machine had the purpose to be able to inform the operator what it has learnt. This will increase capacity, performance and safety. In particular, MAHALO has investigated the impact of transparency (how much the AI is able to explain why it took a specific decision) and conformity (how much the decision taken by the AI is similar to the one a controller would choose).

		AISA	01-06-2020	30-11-2022	Universitat Linz	The project aimed to investigate the effect of distributed human–machine situational awareness in en-route air traffic control operations and also explore the opportunities it entails. To this end, the project was not focused on automating isolated individual tasks but has developed an intelligent situationally aware system.
		ASTAIR	01-09-2023	28-02-2026	Ecole Nationale de l'Aviation Civile (ENAC)	ASTAIR (Auto-Steer Taxi at AIRport) will develop an AI-enabled tool to support a wide variety of ground procedures aimed at optimising resources, and enhancing the safety and efficiency of a wide range of airside operations at the airport. The project will use AI to prompt actions, such as providing clearances to vehicles on the airport aprons and taxiways according to optimal routes and managing fleets of autonomous tugs to further enhance ground capacity. A key focus of ASTAIR will be in optimising humans and AI interaction by tailoring intelligent systems to operators' modus operandi, ensuring logical consistency across manual and automated control.
SESAR 3JU	**Connected and Automated ATM**	**CODA**	01-09-2023	28-02-2026	Deep Blue	The CODA project involves developing a system in which hybrid human-machine teams collaboratively perform tasks. Specifically, the

EUROPEAN PARTNERSHIP
Co-funded by the European Union

						project will show how a system could adapt to specific situations and react accordingly by using advanced adaptable and adaptive automation principles that will dynamically guide the allocation of tasks. The system will assess the operator's cognitive status, use current traffic data to foresee the future tasks that the operator will need to perform in the future, and calculate the impact of those tasks in terms of cognitive complexity. With this information, the system will predict the future mental state of the operator and will act accordingly by developing an adaptive automation strategy.
		HYPERSOLVER	01-06-2023	30-11-2025	Neometsys	The project aims to develop an HyperSolver based on advanced Artificial Intelligence Reinforcement Learning method with continuous reassessment and dynamic updates, i.e. an holistic solver from end-to-end, covering the whole process to manage, density of aircraft, complexity of trajectories, interactions (potential conflict in Dynamic Capacity Balancing timeframe) of trajectories, conflict of trajectories at medium-term and conflict of trajectories at short-term.

		ISLAND	01-06-2023	31-05-2026	Eurocontrol	The project includes the industrial research aimed to create and use airspace capacity, in combination with targeted and effective demand and/or capacity measures. The project exploits the latest advancements in artificial intelligence and machine learning, to supply a variety of supporting toolsets to ATM stakeholders that enable rapid exploration of options for the deployment of capacity-on-demand solutions, whenever and wherever required. The benefits include increased en-route capacity and improved cost-efficiency of ATS provision, without compromising the current safety levels.
	Capacity-on-demand and dynamic airspace	**FASTNet**	01-06-2023	31-05-2026	Indra Sistemas Sa	The project proposes solutions that contribute to the evolution of ATM aviation into an integrated digital ecosystem characterised by distributed data services. It aims at further enhancing the airports and network integration in tactical, pre-tactical and strategic planning through the development of two solutions: -Enhanced AOPs-NOP Tactical planning, with the inclusion of an "airport-to-airport(s)"" AOP to AOP collaborative planning process and the use of artificial intelligence.

						- AOP-NOP Strategic and pre-tactical planning.
		KAIROS	01-06-2023	31-05-2026	Intelmet Solutions	KAIROS will improve the quality of meteorological information provided to the aviation community through the use of artificial intelligence. By producing accurate digital weather forecasts at longer lead times, aviation stakeholders will be in a better position to mitigate the impacts of weather on their operations. The project will integrate live weather information from AI forecasts with existing decision support tools and platforms to assess the operational benefits to several end-users.
		AI4HyDrop	01-09-2023	28-02-2026	Universitet et i Sorost-Norge (USN)	With an increasing number and diversity of potential drone operations, managing the airspace to accommodate these drones will become an increasingly sophisticated task, especially in densely populated urban areas encompassing restricted zones with dynamic environmental and operational influences. Due to the associated higher probability of conflicts, and ultimately collisions, such areas require management of dedicated structured airspace, operations, and services to help mitigate these potential hazards. A holistic framework is necessary to create an

EUROPEAN PARTNERSHIP

Co-funded by the European Union

						effective and efficient flow of information between the various capabilities in order to systematically organise the airspace usage. AI4HyDrop evaluates the various stakeholder needs, delivering validated concepts, defining a methodology for an airspace structure organisation and associated U-space services. The framework considers the information from other services such as meteorological and separation provision, which can then be used for flight planning approval process, prioritisation. In addition, essential elements such as surveillance and contingency planning can be addressed. The framework incorporates various AI based tools and associated information flows necessary to address the complexity, safety and scalability required for implementing such U-space services.
	Aviation green deal	**SynthAIR**	01-09-2023	28-02-2026	Sintef	The project aims to respond to the scarcity of relevant data for aviation and the inherent limitations of AI models in handling diverse datasets. The main idea is to learn a model from multiple datasets and generate synthetic data that accurately represents new, unseen datasets, through the groundbreaking concept of the Universal Time Series Generator (UTG).

	Artificial Intelligence for aviation	**TRUSTY**	01-09-2023	28-02-2026	Malardalens Universitet	Remote digital towers represent one of the last innovations in aviation, offering remote traffic flow and capacity management for airports. While conventional control tower host operators have direct visual oversight of runways and taxiways, digital towers exploit video transmission to provide the same vital information. This advancement enables the provision of airport air traffic services (ATS) from virtually anywhere, promising significant enhancements in operational efficiency and safety by augmenting controller situational awareness. In today's era, artificial intelligence and machine learning are offering automated and faster solutions in many industries, bringing the industry to a more advanced stage.
		ASTRA	01-09-2023	28-02-2026	Universitata Malta	Nowadays, tactical Air Traffic Control (ATC) hotspots are only identified up to around 20 minutes in advance. The aim of ASTRA is to bridge the gap between the Flow Management Position (FMP) and the planner Controller Working Position (CWP) by developing a AI-based tool for FMP personnel which can predict and resolve hotspots earlier than today, before they are within the scope of the sector planner. The objectives of the project

						are to: develop an FMP function to predict hotspots at least 1 hour in advance, and to propose strategies to resolve them; develop Human Machine Interface (HMI) concepts to allow interaction between operators and the tool; and demonstrate and validate the tool by conducting human-in-the-loop Real-Time Simulations (RTS) in a representative operational environment.
		JARVIS	01-06-2023	31-05-2026	Collins Aerospace Ireland	The project addresses the increasing complexity of the entire aviation ecosystem (aircraft, air traffic control – ATC, airports), through the introduction of a Digital Assistant (DA) that, by teaming with its human counterpart (pilots, ATC operators, airport operators), support the execution of tasks to ensure safe and profitable operations in complex scenarios. JARVIS Consortium aims at developing and validating three ATM solutions: an Airborne DA (AIR-DA), an ATC-DA and an Airport DA (AP-DA). The AIR-DA will increase the level of automation in the flight deck and thanks to AI-based actions will act as enabler towards reduced crew operations and single pilot operations. The adoption of the AIR-DA will allow pilots to deal with complex scenarios without

EUROPEAN PARTNERSHIP
Co-funded by the European Union

						compromising safety, security, while reducing the pilot workload. The ATC-DA will increase the level of automation in control towers, where environmental KPIs and the capacity management of airspace will benefit from the adoption of AI-based technologies. Finally, the AP-DA will increase the level of automation in airports, enhancing safety and security for intrusion detection scenarios.
		DARWIN	01-06-2023	31-05-2026	Honeywell International	DARWIN's ambition and vision is to develop technology enabling AI based automation for cockpit and flight operation as a key enabler for SPO (Single Pilot Operations) and demonstrate the same (or higher) level of safety with same (or lower) workload as operations with a full crew. The system will consist of 3 core enabling technology layers: 1) Trustworthy Machine Reasoning Platform will provide capabilities for rule-driven, transparent, and explainable decision aiding or decision making. 2) Human-AI Collaboration layer will be implemented on top of the Reasoning Platform. 3) Pilot State and Taskload Monitor will provide data to the collaboration layer and automation to adaptively react.

Horizon		**HAIKU**	2022	2025	Deep blue	Human AI teaming Knowledge and Understanding for aviation safety – European-funded project aimed at enhancing Human-AI teaming for future aviation systems in the 2030C timeframe (https://www.haikuproject.eu/). HAIKU has six human-centric AI use cases, two each in the air traffic, cockpit and airport sectors, where prototype Digital Assistants will be developed. HAIKU aims to explore human-AI interactions and teaming in dynamic and realistic simulations of operational flight scenarios. Three main research questions will be addressed: • What is the recommended human-AI relationship for each of the different AI applications in aviation? • What does it mean for AI to be explainable and hence trustworthy in each of these applications? • How do we best teach AIs, via human-in-the-loop AI learning for each of the potential aviation applications? The following main outputs are foreseen: 1. New Human Factors design guidance and methods ('HF4AI' Capabilities) on how to develop safe, effective and trustworthy Digital Assistants for Aviation 2. A set of aviation use cases – controlled experiments with high operational

EUROPEAN PARTNERSHIP
Co-funded by the European Union

						relevance – illustrating the tasks, roles, autonomy and team performance of the Digital Assistant in a range of normal and emergency scenarios 3. New safety and validation assurance methods for Digital Assistants, to facilitate early integration into aviation systems by aviation stakeholders and regulatory authorities 4. Continuous engagement with relevant stakeholders - e.g. policy makers, professional associations, passengers associations and general public – to deliver Guidance on socially acceptable AI in safety critical operations, and for maintaining aviation's strong safety culture record.
		SafeTeam	1 July 2022	30 June 2025	Fundacion instituto de investigaci on innaxis	Safe Transition to Digital Assistants for Aviation - The goal of the project SafeTeam is to progress on the human factors aspects of the use of digital assistants to aviation, including a deeper understanding on the technology and processes that will facilitate the adoption of AI tools and integration into operations, enhancing human cognitive abilities and potentially automation. SafeTeam is not purely concerned with the technical development of AI applications for aviation but rather focuses on those aspects and characteristics of integrating such digital assistance / AI tools that will ensure efficient and safe

						interaction with human operators. The efficiency / accuracy of the ML algorithms and AI solutions is of course of relevance to the research, but the core objective of SafeTeam is to facilitate the transition to digital assistants and ultimately AI-run operations from a Human Factors and safety perspective. The project will also look into approval and certification issues, concretely on aspects related to the human ability to operate sophisticated AI tools and explainability of AI operations. The project's main goal of developing new human-machine interaction concepts will run along important technical challenges required to reach TRL6, demonstrating several concrete use cases in relevant environments placing human operators at the core of the research. Particularly, leveraging the past work done on data infrastructures, the SafeTeam project will provide relevant environments integrated with the use cases, to be able to demonstrate the different ML algorithms and the human interactions with the enhancing awareness or automation case studies presented.
		Digital Assistant: Introducing	2018	2020	Civil Aviation Authority	This year at the Singapore Airshow, AIR Lab unveiled its latest Proof-of-Concept that aims to revolutionise the

EUROPEAN PARTNERSHIP
Co-funded by the European Union

		Speech Recognition			of Singapore	working ways of air traffic controllers – the Digital Assistant. Developed in close collaboration with Civil Aviation Authority of Singapore (CAAS), Thales and the Agency for Science, Technology and Research (A*STAR), Digital Assistant leverages speech recognition technology to automate processes and streamline operations, eliminating the need for manual data entry of commands.

4.6.2 Opportunities and Challenges

Opportunities

1) The application of new technologies and interdisciplinary research brought by the introduction of the human-centric concept and HMI design focuses on sustainability and resilience.
2) Improvement of workers' well-being.
3) Improve the relationship between human and machine collaboration.
4) Assure the best of human and the best of machine joint working.

Challenges

Considering the opportunities, the following challenges may become an important direction for research:

1) **Human AI Teaming** (National Academies of Sciences, Engineering, and Medicine, 2022)
    a. Existing human-AI research is severely limited in terms of the conceptualizations of functions, metrics, and performance-process outcomes associated with dynamically evolving, distributed, and adaptive collaborative tasks. Research programs that focus primarily on the independent performance of AI systems generally fail to consider the functionality that AI must provide within the context of dynamic, adaptive, and collaborative teams. Research should specifically consider the dynamic process factors and timing constraints involved when human-AI team members address uncertainties in task progress or the evolution of performance over work sessions, shifts, task episodes, software updates, and longer time horizons.
    b. Human-AI Team Models. Predictive models of human-AI performance are needed to provide quantitative predictions of operator performance and interaction in both routine and failure conditions.
2) **Performance Framework** to assess the effectiveness of the “assistance” and the potential integration with SESAR Performance Framework
3) **Technologies of multimodal fusion perception and human-like intelligent perception.** The perception of the human emotional status is relevant for a good human-machine interaction. It is a kind of empathy to be implemented at machine level and to tune as a consequence its behaviour. Human beings express their emotions and intentions through multiple signals, such as language, pronunciation, and intonation, facial expressions and gestures, as well as some physiological signals, such as blood pressure and heartbeat. Most of the existing perception methods are focused on the single mode. The correlation between the multiple modes is ignored. Therefore, the creation of multimodal databases, multimodal data hierarchical fusion perception, and human-like intelligent perception technologies based on this database (R 11) can be explored.
4) **Mechanism of multimodal cooperative analysis and intelligent reasoning**. At present, intelligent reasoning leverages on algorithms conditioned in some way by the status of the human. Once the multimodal perception is explored, a deep adaptive cooperative semantic understanding mechanism is needed based on ontology.

Co-funded by the European Union

5) **Deep understanding of natural language and personalised interaction.** In the personalised interaction, the intelligent robot can adjust the interaction method and strategize neatly according to the scene, interaction object, interaction state, etc.

6) **Personalised interaction to cope with biases of the human- a new form of training.** Once the assistant is "profiled" on the assisted human, typical human biases in specific conditions should be detected and known by the Assistant that can avoid the error, after which the human will need to be able to understand the status of the machine and cope with its biases.

7) **Effective computing**. By using the abilities of perception, deduction, and prediction, intelligent robots or computers are involved in a large number of tasks in our daily life. The key issue is that these robots are not similar to humans from the perspective of emotions. It is well known that emotion is a necessary factor for communication and interaction between humans. Therefore, people naturally expect intelligent robots to have EQ along with IQ. *Can it be an enabler for an effective man-machine teaming?*

8) **Integration of human and machine**. As the level of intelligence grows and the human overlies on it more and more, the task allocation and the workload have to be further investigated.

Regarding the XAI

Challenges opportunities

The market of XAI is expected to grow with 20.9% in the next five years representing a great opportunity.

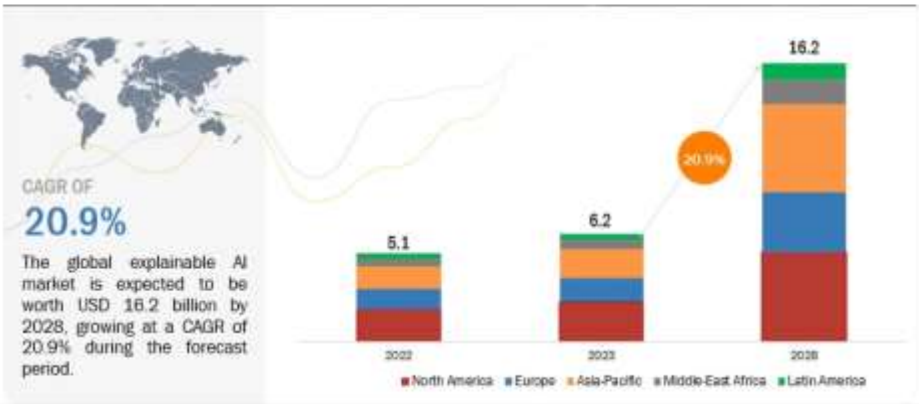


Figure 34 Explainable AI market forecast[3]

(Saeed, W. et al., 2023) discusses the challenges and research directions of XAI in the deployment phase (see Figure 16).

---

[3] https://www.marketsandmarkets.com/Market-Reports/explainable-ai-market-47650132.html?gad_source=1&gclid=CjwKCAiA-bmsBhAGEiwAoaQNmqFQPK_X2bq34rchI3QUxg8zNUiitKOfaekTC7_HUxjvO7M32XaevxoCNFIQAvD_BwE

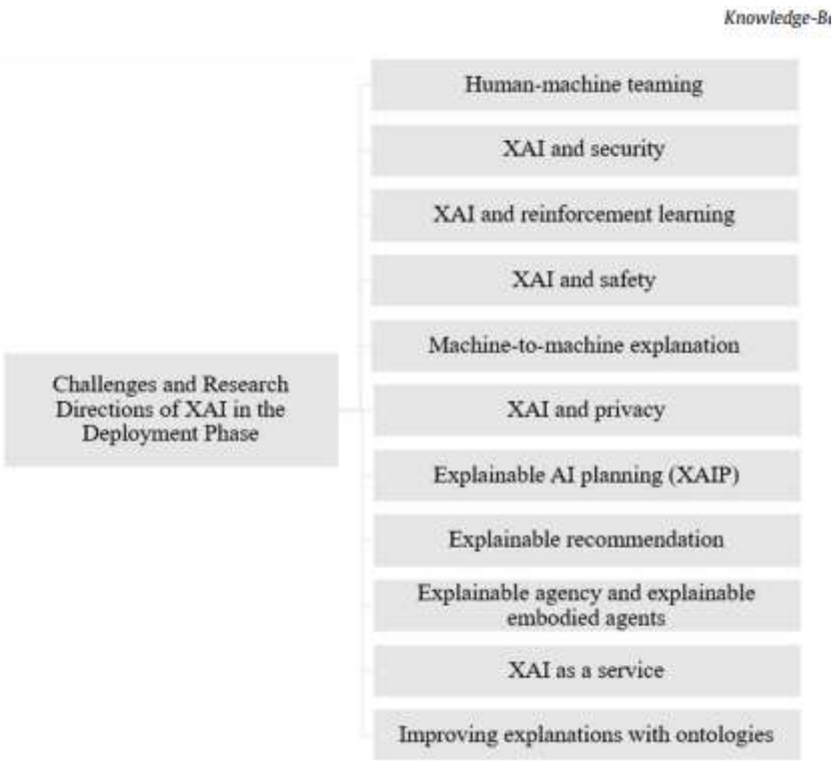


Figure 35: Challenges and Research Directions of XAI in Deployment Phase (Saeed, W. et al., 2023)

4.6.3 Conclusion

The present work on Human Assistant has provided the most recent current developments on Human Assistant giving some elements of current research on key aspects of Human Assistant: Man-Machine interaction, Human Ai Teaming, Trust, Explainability, Task Allocation and Performance metrics. Human assistant is a topic transversal to all domains and the need of a unique theoretical framework for taxonomy, research proper classification and benchmark is needed.

For these purposes, the framework proposed in Industry 4.0 and 5.0 could be adopted creating a great potential of cross fertilisation among the different sectors once they are speaking the same language. For example, during such study we found that task allocation problems are faced according to certain paradigms when talking of digital Agents and very extended studies in the robot domain (Multi Robot Task Allocation). Many aspects of MRTA may apply to digital agents.

Furthermore, the metrics to assess the performance of an Assistant could be a starting point to set up a link with the SESAR framework expected to consider in the next future the value of the assistant ecosystem.

Accordingly, the present work provides a very brief overview of the huge world behind the concept of Human Assistant not aiming at being exhaustive but instead aiming at triggering useful questions on the basis of the opportunities that AI Assistants can rise and the related challenges.

5 Opportunities and Challenges

5.1 Workshop

In Work Package 2 (WP 2.3), our primary objective was to analyse the impact of automated systems and AI on aviation. We focused on identifying key opportunities and challenges, with a particular emphasis on the human factor and safety concerns. To accomplish this goal, a dedicated workshop session was organised to collect and discuss "Opportunities and Challenges." The virtual event was held on January 29th, 2024, and utilised the "Miro interactive whiteboard" to facilitate active participation and feedback. The workshop was attended by 11 participants from the entire HUCAN consortium, including partners who were not involved in WP 2. The session spanned from 09:00 to 14:00 and consisted of three parts.

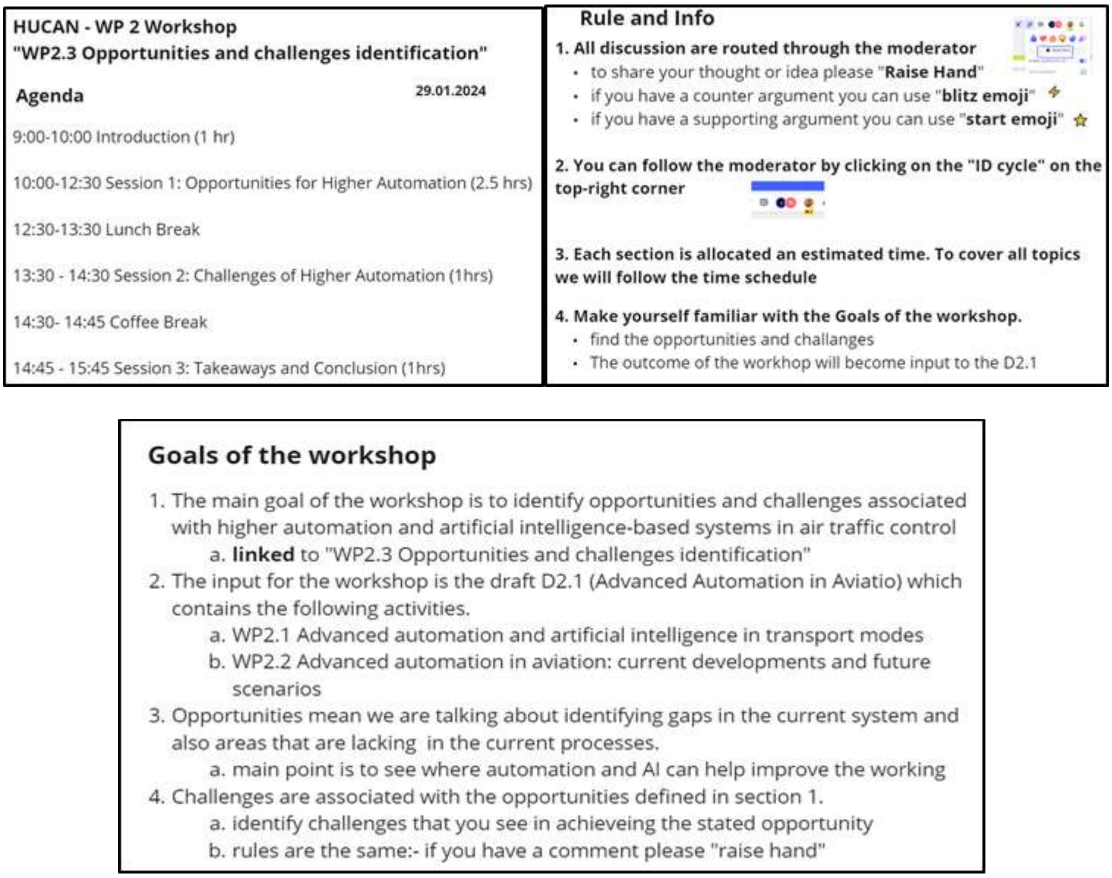


Figure 35 WP 2 workshop schedule and goals

To generate ideas and gather thoughts from a diverse group of experts, we chose the brainstorming workshop method. To keep the discussion organised and extract the most information, we communicated a set of rules to all participants. These rules included registering and expressing opinions in favour or against a point. We also moderated two sessions to ensure that each point was adequately understood and discussed. Additionally, we established general rules such as "the more the better," "there is no such thing as a bad idea," and "think in terms of: what if" for open discussion.

Co-funded by the European Union

Throughout the workshop, participants were actively involved in a discussion surrounding the complexities and venues of automation in aviation. To facilitate an effective discussion of broader views, the workshop was structured into four distinct sessions, each of which had a specific purpose.

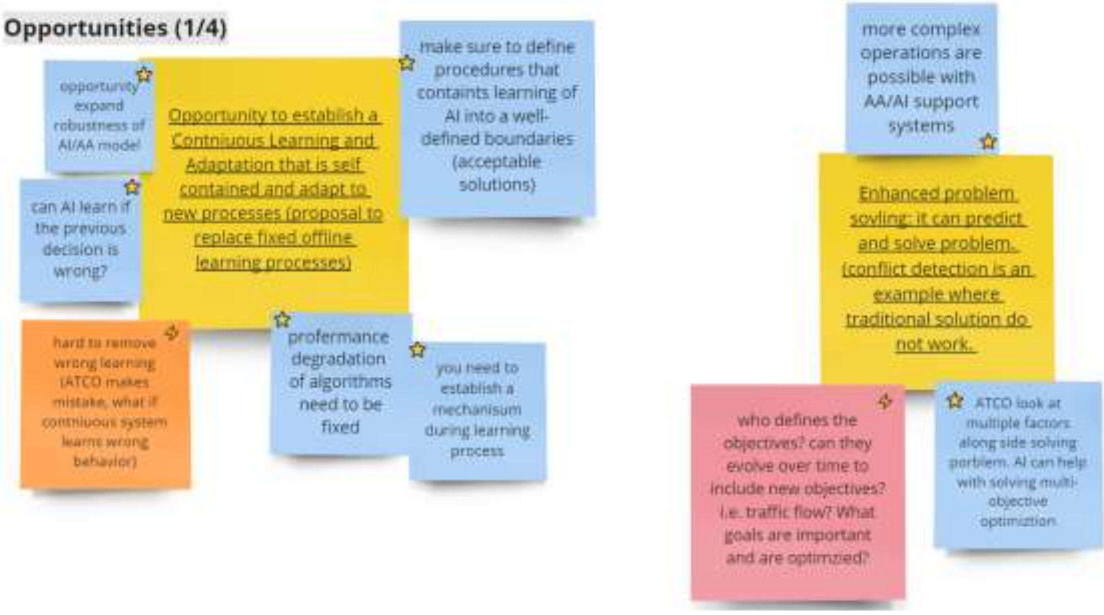


Figure 36 Opportunities (1/4) collected and discussed during the HUCAN WP 2 workshop

The first session, which was the introduction and idea generation session, set the tone for the rest of the workshop. In the introduction, participants became familiar with the workshop's objectives and principles, and the ground rules were established to ensure an inclusive discussion. After that, the remainder of the time was spent gathering perspectives about the current state of automation, the identification of gaps and the identification of future opportunities, allowing everyone to align to the main topic and focus on assessing the current landscape and identifying areas for improvement. The opportunities identified in this session laid the groundwork for the subsequent sessions.

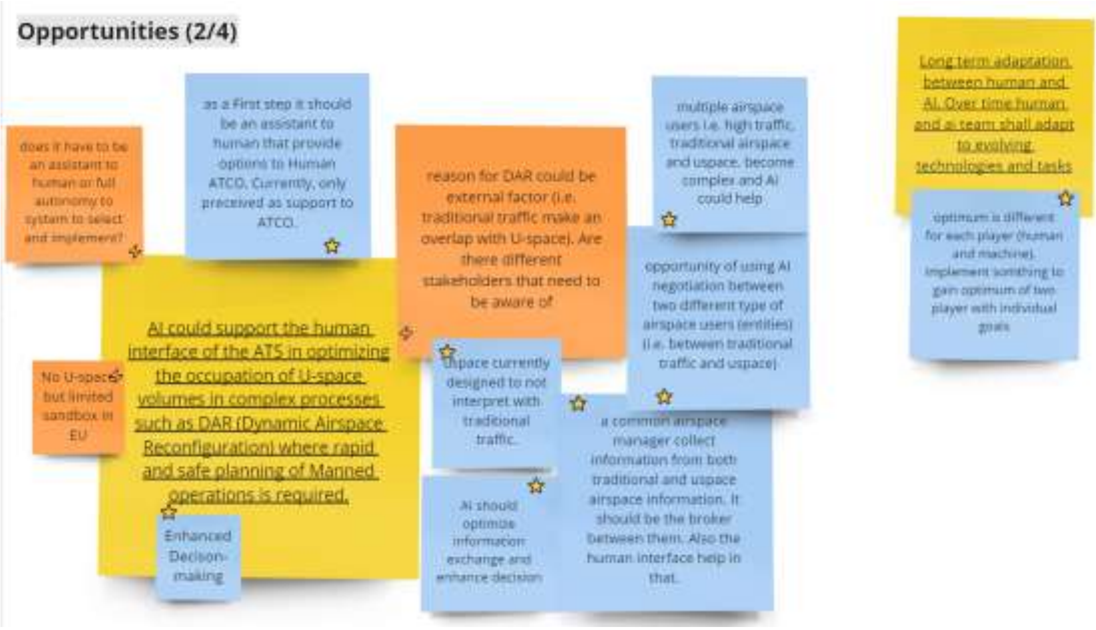


Figure 37 Opportunities (2/4) collected and discussed during the HUCAN WP 2 workshop

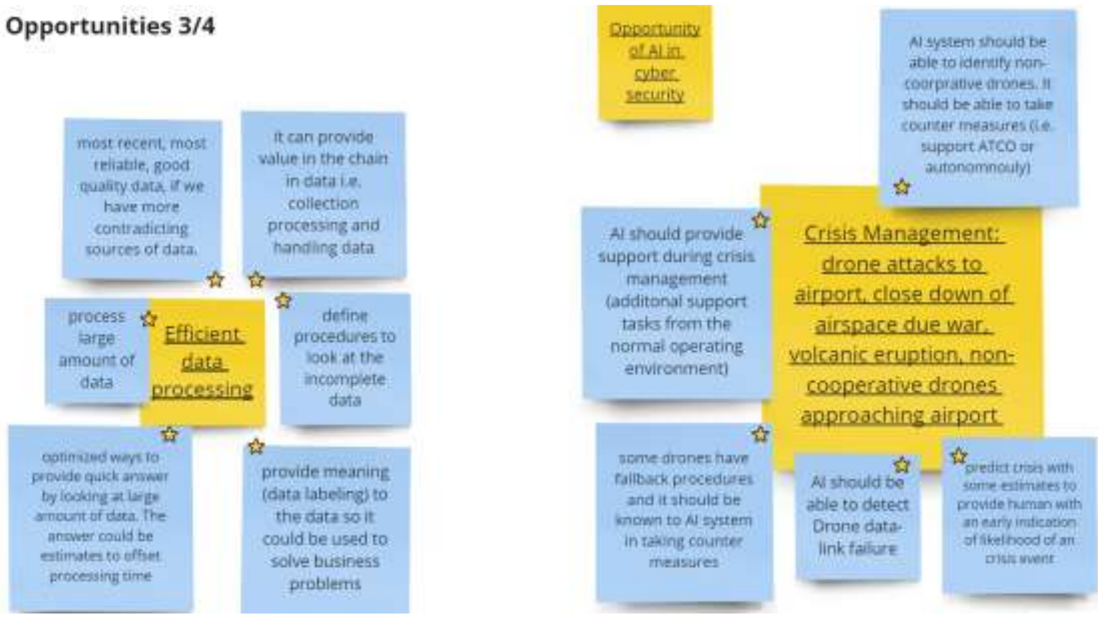


Figure 38 Opportunities (3/4) collected and discussed during the HUCAN WP 2 workshop

During the second session of the workshop, participants took part in the opportunity exploration and evaluation phase. Here, they examined the opportunities that were discovered in the previous session. Through detailed analysis and in-depth discussions, the participants were able to gain a comprehensive understanding of each prospect. They shared valuable feedback and insights, which helped to identify any potential opportunities. The discussions were lively and engaging, with participants critically examining each idea and exploring all possible angles. By working together, they were able to collectively assess the feasibility and potential impact of each opportunity.

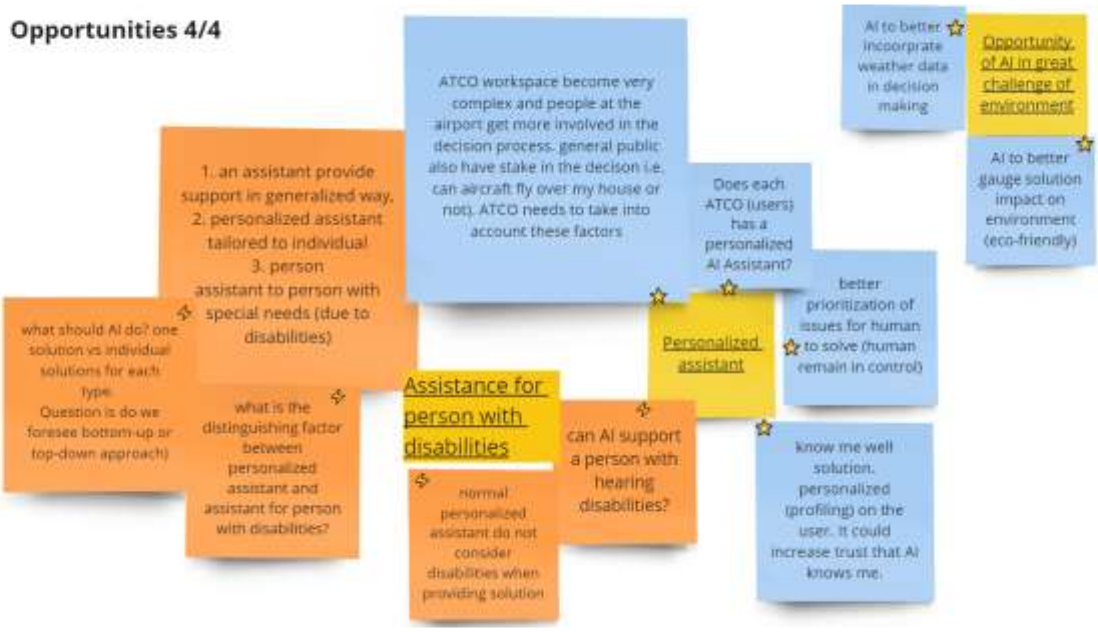


Figure 39 Opportunities (4/4) collected and discussed during the HUCAN WP 2 workshop

Towards the end of the session, participants were given a voting scale which included "Do it Now", "Do it Next", "Do it if we have time" and "Don't Do it". This scale represented the priority in implementing each opportunity. They used this voting scale to assign priority to each opportunity. At the end of this session, the total votes were counted and the opportunities that were deemed "Do it Now" were chosen for the next round.

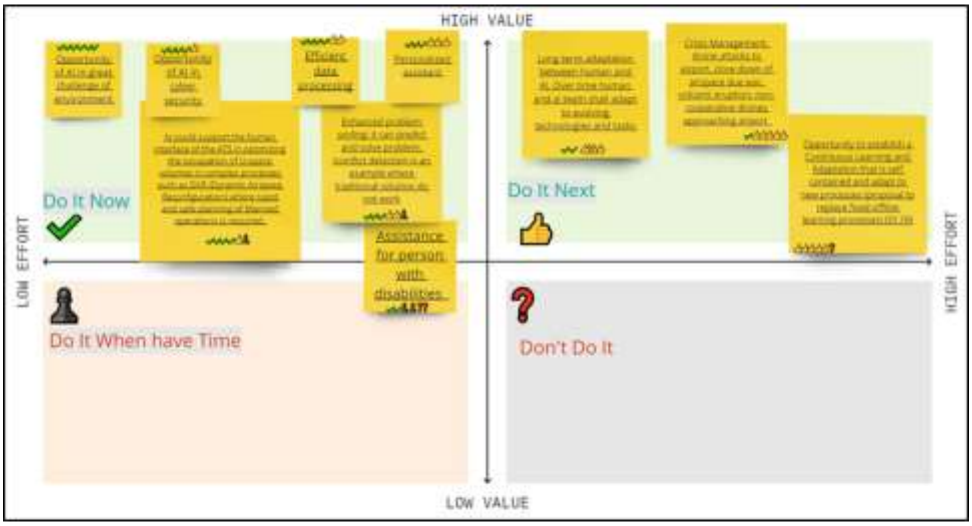


Figure 40 Classification of opportunities into four quadrants according to importance generated during the HUCAN WP 2 workshop

During the third session of the workshop, participants transitioned from exploring ideas to creating actionable plans. They shifted their focus towards a more practical approach, identifying the opportunities that were most urgent and needed immediate attention. With a keen eye for detail, they carefully scrutinised each opportunity, looking for any potential roadblocks or challenges that could hinder their progress. Through candid and strategic deliberation, participants proactively sought to mitigate risks and devise actionable strategies for implementation. They engaged in open and honest discourse, sharing their insights and perspectives to ensure that the best ideas were brought to the table. By working collaboratively, they were able to overcome obstacles and devise practical solutions that could be implemented effectively.

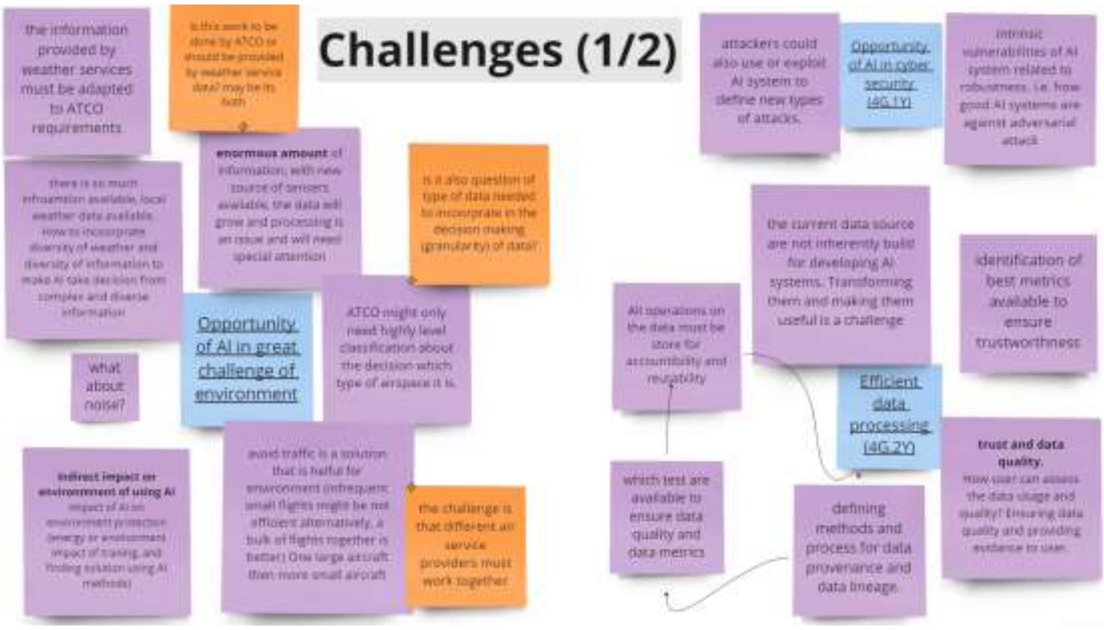


Figure 41 Collection of challenges (1/2) associated with opportunities discussed during the HUCAN WP 2 workshop

EUROPEAN PARTNERSHIP

Co-funded by the European Union

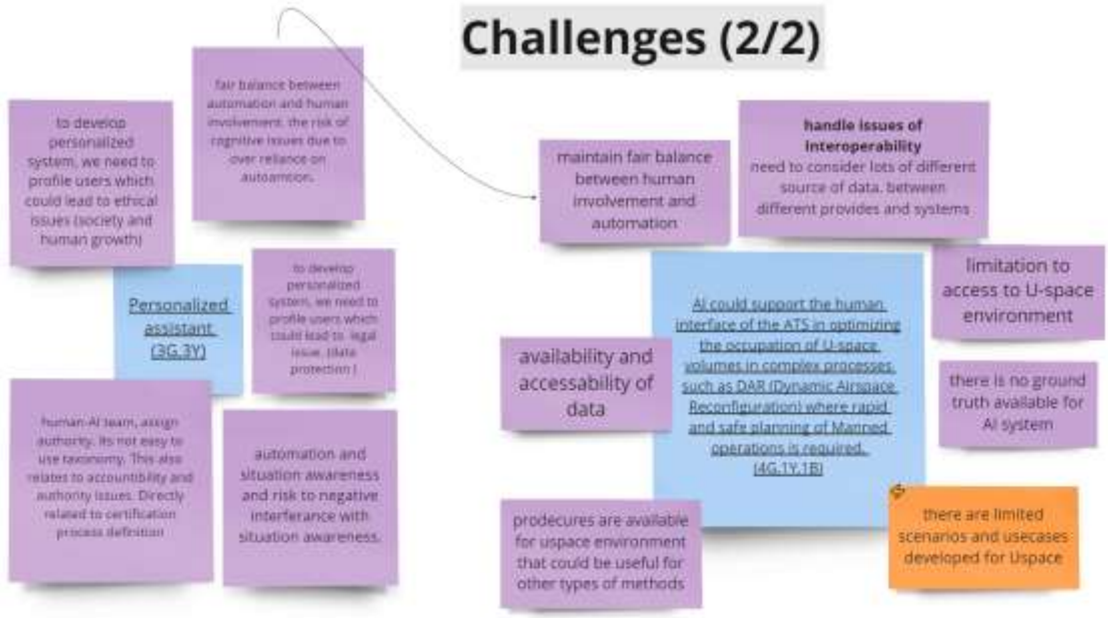


Figure 42 Collection of challenges (2/2) associated with opportunities discussed during the HUCAN WP 2 workshop

The last session was dedicated to a final discussion about synthesising the main takeaways and determining what steps to take next for WP 2. During the workshop, participants identified 11 opportunities and examined each in detail. Five of those 11 opportunities were shortlisted, and potential challenges were identified. Overall, the workshop was a great success as it generated intense and productive discussions in a concise amount of time. As a result, a collection of opportunities and challenges were gathered, focusing on higher automation in aviation. For a detailed description of each opportunity and challenges discussed during the workshop, please refer to the following two subsections.

5.2 Opportunities and Challenges

AI offers opportunities through

1. Its possibilities to quickly process large amounts of data from different sources, even if the data is conflicting, incomplete or inaccurate;
2. Its characteristics to learn, adapt and predict;
3. Its ability to flexibly support tasks of human operators through building task models and profiling;
4. Finding new solutions to issues that cannot be solved with conventional means, like conflict detection and solutions to “larger-than-aviation” problems (e.g. general world-wide sustainability).

The following subsections discuss opportunities and corresponding challenges for the use of AI in aviation, which were collected during the workshop.

5.2.1 Efficient data processing

- Opportunities

AI can process complex, non-accurate information. Most AI systems are capable of processing large amounts of data; even if the data originates from different sources and contains conflicting information, it can still be used by the algorithms. Just as well, AI is capable of reasoning with incomplete data or with information that is processed and provided as an estimate, thus containing uncertainty.

This efficient data processing is a major benefit of the use of AI/AA in complex decision-making tasks.

- Challenges

The reasoning process and the solution presented to the human operator must be clear, so that a decision is made based on well-motivated grounds. The information presented must be meaningful to the human decision-maker so as to solve business problems and provide value in the chain of collecting data, information processing and decision-making.

The technical challenge of the use of AI is that it needs to be decided what AI techniques to use for specific situations in which data/information needs to be processed. Information from different sources that conflicts, uncertain information, incomplete information or information that contains estimates requires careful analysis before a decision is taken on what AI technique to use.

5.2.2 Continuous learning and adaptation (self-learning systems)

- Opportunities

Self-learning systems offer the possibility to adapt continuously while in operation. This can be a powerful feature in complex environments in situations where errors have been introduced in the process. The system will thus autonomously correct errors.

The system is also flexible in learning new things from entering into new operational modes and new operational environments. Just as well, new systems can be introduced in the aircraft or on the ground, where, through a self-learning adaptive process, the newly introduced elements will be taken into account without the need to change the software that processes the new information.

The possibility to expand the AI/AA-model offers a robust environment. Instead of off-line learning and the need to install new systems with every system upgrade or new operational environment, the system now is capable of adapting.

- Challenges

An unsupervised learning process can be applied without risk for non-critical system elements, however, for decisions that impact safety, it must be ensured that the newly adapted system will not cause any harm.

Learning during the process of operation implies that all aircraft will learn differently and will have other systems as they have been operating in other environments.

One major challenge is that a mechanism must be in place to learn. As structural changes to e.g. the airspace are implemented, the trained capabilities of a support system become obsolete. After such changes, new training data must first be collected, processed, and thus taught to a system in order to be able to provide support in the new environment. This extension to the system does not have an operational function in the existing software but is an internal processing system that is modifying answers of systems that have been approved in the past already. This kind of software is new to aviation, where currently all software must be in direct function to an operational process.

Procedures must be in place that ensure the AI is learning within well-defined boundaries and creates safe and acceptable solutions. The self-learning system must guarantee the learning process to “move” towards operationally acceptable solutions and should not get stuck into some loop of non-viable solutions.

5.2.3 Personalised assistants

- Opportunities

Each individual will have a dedicated way of working, even though many of the tasks in aviation have been standardised. With increasingly complex work, a personal touch will be given to each task. Through personalised profiling, the user can be assisted by AI systems to perform their tasks without having to slightly divert from their preferred way of doing so. Central to this adaptation is the understanding that the ideal outcome varies for each participant involved in the human-AI collaboration.

Through task models the AI system can propose priorities to the human operator and assist in scheduling tasks.

Another opportunity of the use of AI in task modelling is that the work can be checked and small personal mistakes or errors can be eliminated from the process as quickly as possible.

- Challenges

The construction of task models for human operators is a complex process. The task must be detailed at the correct level and the relation between tasks must be carefully laid down, allowing an understanding of the consequences of carrying out (part of) a task later in the process or reversing the order of tasks. Pre- and postconditions must be clear and the context of the task must be considered.

User profiling requires a longer assessment of the user, who may even change their behaviour in the course of the process. The development of a personalised system could lead to ethical and legal issues.

To assist the user, based on their profile, a sequence of tasks can be offered, that proposes the preferred task to be performed, where care must be taken that its priority must be in line with the required objective of the work of the human.

Small errors must be detected before they become serious, however these should not be presented as annoying. The human operator might have left the task out at the moment to give personal preferences to a task or consider the task to be of lower priority at the moment.

5.2.4 Long-term adaptation between humans and AI

- Opportunity

As the role of the human is continuously changing while AI technology continues to advance, so too must the relationship between humans and their artificially intelligent counterparts. With growing experience, the human will consider and perform certain tasks differently than when operating as a novice. Just as well, the tasks may change over time with new systems and technology, new procedures or a new operational concept and the AI support system can grow with its human operator. Particular attention is given to the design of the human-machine interface (HMI) to facilitate seamless collaboration in AI-assisted decision-making.

In cooperating with AI systems, humans will perceive the role of both human and machine as changing over time. The continuous support and experience gained from working with an AI system will allow the human to consider the system as more reliable and trustworthy as long as the support given is correct. The human might even consider giving certain tasks to the system that were initially performed manually leading to enhanced productivity, efficiency, and innovation.

This could include providing opportunities for skill development and knowledge acquisition for human operators, as well as enabling AI systems to continually refine their algorithms and adapt to new challenges.

- Challenges

To allow cooperation between human and machine, an intuitive and clear user interface is required that enables the full potential of the partnership. The interface must be able to address complex problems in aviation but also in human-machine interaction.

The relation between the human and the AI system must be carefully mapped to allow such complex cooperation. Humans may even expect a changing role from advanced automation when they learn to master their tasks better.

5.2.5 Support in crisis management

- Opportunities

In situations of crisis-management, AI can support the human operator by taking over the standard tasks that an operator has, but which are now given less priority because of the crisis.

With good prediction systems, crisis situations can be discovered beforehand. Crises require preparedness at any moment and the sooner the situation is recognised, the better-prepared decisions can be taken. AI can especially contribute to surveillance and mapping all aspects of the crisis, providing situational awareness to responders.

- Challenges

Find a new division of tasks between the human and the system, in accordance with the crisis situation at hand and acceptable workload for the human.

Preparedness for crisis situations asks for a systematic approach towards crisis types and preparedness in different responses. Timing is crucial.

5.2.6 Enhanced problem-solving

- Opportunities

AI offers enhanced problem-solving for issues that at the moment cannot be solved, so that more complex operations become possible, e.g. with the use of drones.

AI/AA offers new techniques beyond current problem-solving algorithms that are capable of extremely fast processing of data and information. The logic processes behind the algorithms provide a kind of intelligence that enhances traditional algorithms. This feature allows new operational problems to be solved, like conflict detection and planning of large numbers of drones in complex (urban) environments.

In many complex situations, the objectives of solving an issue are not all in line with each other, requiring a multi-objective problem-solving technique that considers optimal solutions instead of always finding the one "best" solution. It may even be that the solution proposed does not rank highest in the list of optimum solutions, but provides acceptable and safe solutions in the given time for finding a solution.

A special enhanced problem-solving feature is the prediction that AI systems can deal with the uncertainty of future situations. This may aid for example the above-mentioned conflict detection and just as well offers new opportunities in special operational situations. AI/AA offers prediction techniques that may enhance the safety of the whole aviation system.

- Challenges

Instead of considering the safety of an algorithm, in a complex environment, the safety of the objective function should be considered. The system will propose solutions, within limited time, that are considered "optimum" solutions to new complex challenges.

In enhanced problem-solving, the AI system will need to deal with multiple objectives to find a solution. These objectives may, at some operational level, even be conflicting with each other. In finding an optimum, it may be possible to end the process of finding an even better solution through presenting a local optimum. The question becomes who defines the objectives, especially when they evolve over time and include new objectives. Further, who decides the relative importance of each of the objectives?

Dealing with uncertainty, especially in predictive behaviour is a characteristic that requires reasoning with uncertainty. This is a dedicated field in AI with new challenges such as what would be the uncertainty in situations where decisions are taken, based on this prediction?

5.2.7 Dynamic Airspace Reconfiguration

- Opportunities

AI can support the Dynamic Airspace Reconfiguration (DAR) process by adjusting U-space airspace safely and efficiently to allow manned aviation to pass through while still maintaining an optimum volume of airspace to accommodate unmanned traffic. The opportunities mentioned in this section concern human support in the reconfiguration process.

The first opportunity is to provide the human decision-maker (e.g. the ATCO or a dedicated human broker) with options to support the reconfiguration process. For each possible option a motivation can be provided, e.g. in terms of timing, volume of airspace required and the number of manned and unmanned vehicles that can be accommodated in the solution.

Another opportunity for an AI system is to provide the means to negotiate between the different types of aircraft that plan to use the same airspace. Negotiation is an iterative process requiring negotiation parties to work on proposals and counter-proposals to reach a solution that is agreeable for both. This often means one might have to extend some concession to get favour in return.

- Challenges

Explanation of reasoning in optimisation problems is a challenging task. The different objectives of the optimization problem can be given a (numerical) score and then explained to the human operator, though this is usually not according to the terminology he would use. Furthermore, it will be difficult to make a good assessment of the value of each of the objectives and compare these with the values of others.

Negotiation is a separate AI topic that still requires more research. Negotiation takes place in a larger context, possibly extending the scope of the actual topic of the negotiation. Just as well, the process requires more than one instance over time. An issue with a solution that was considered yesterday to be unfavourable for one user might be solved differently to give that user the benefit some other time.

5.2.8 AI and the greening of aviation

- Opportunities

Reducing climate-damaging influences, such as emissions of CO2, NO_x, water vapour and condensation trails, is a challenge also for air traffic and AI-based innovations can help to find solutions that meet sustainable goals. The analysis of collected data regarding emissions could be translated into patterns by machine learning and thus into more accurate estimation of environmental impact which is essential for generating green trajectories.

- Challenges

The generation of green trajectories is a complex computational problem that needs to consider a large number of parameters, including the still unpredictable weather patterns. The availability of information is a challenge. Weather phenomena are diverse in terms of complexity, type, duration, and variability and can occur very differently locally. In order to integrate these into an AI system, extensive training data is required, which must cover a wide range of weather events as well as traffic situations.

Since AI itself is part of the solution, it should not be neglected that the use of AI also consumes resources for computing power, storage and cooling. Renewable energy can reduce the environmental burden, though measuring the environmental impact of AI in computing and its applications is currently limited by the lack of recognized standards, consistent indicators and metrics.

5.2.9 Conclusion

The following table summarises the above-mentioned opportunities and challenges.

Table 8 Summary of Opportunities and Challenges

Efficient data processing		
Opportunity	**Challenge**	**Issue**
Support human decision making	Provide information in meaningful elements to support the chain from data to information to decision	Human operators will not be able to understand the AI reasoning process if this contains merely figures to compare
Reasoning with data from different sources, incomplete data, uncertain data or estimates	Find the right AI-technique to process the data	
Continuous learning and adaptation		
Opportunity	**Challenge**	**Issue**
Correct error	Do not get stuck in loops	Before errors can be corrected, they will be made. This can be a safety issue
Adapt to new environment	Ensure viable and safe solutions at all times	In many situations, the system will not be allowed to learn through making mistakes
Adapt to new systems (1)	Ensure viable and safe solutions at all times	In many situations, the system will not be allowed to learn through making mistakes
Adapt to new systems (2)	Operationally non-functional software must be installed	
Adapt to new systems (3)	All aircraft will have their “own” system different from others	
Personalised assistant		
Opportunity	**Challenge**	**Issue**
Support human tasks (1)	Build human task models	

Support human tasks (2)	Profiling requires advanced algorithms to understand human behaviour	
Support human tasks (3)	Profiling could lead to ethical and legal issues	
Prioritise human tasks	Understand links between tasks in task models	
Eliminate human errors as early as possible	Understand links and consequences in task models	
Long term adaptation between humans and AI		
Opportunity	**Challenge**	**Issue**
Support humans in an environment that changes over time	Mapping of human-machine cooperation necessary	
Changing relationship between humans and their AI support systems	Design the human-machine interface carefully so that an intuitive cooperation comes to place	With learning to control their tasks, humans will expect a changing role from advanced automation
Human skill development through cooperation with AI systems	Mapping of human-machine cooperation necessary	
Support in crisis management		
Opportunity	**Challenge**	**Issue**
Support human in crisis situations by taking over part of the routine job	Find a good division of tasks between human and AI, according to the situation at hand an acceptable workload for the human	
Support in surveillance and providing a quick overview of the crisis situation	Timing and being prepared for crisis situations asks for a systematic approach towards crisis types	
Enhanced problem solving		
Opportunity	**Challenge**	**Issue**

Solve complex problems (1)	Find multi-objective solutions	
Solve complex problems (2)	Dealing with local optimum and stop searching for "better" solutions	
Multi-objective optimisation (1)	How to deal with new objectives. Who defines them?	
Multi-objective optimisation (2)	Who decides the priority of each objective	
Dealing with uncertainty (providing e.g. predictions)	How to make decisions on uncertain information	Processing of uncertain information
Dynamic Airspace Reconfiguration		
Opportunity	**Challenge**	**Issue**
Support human decision making	Explain solutions obtained through multi-objective optimisation	Comparison between objectives will be difficult to make for the human
Support the negotiation process (1)	This is an AI topic that requires further research	
Support the negotiation process (2)	Negotiation takes place in the context of a larger environment	An issue with a solution that was considered yesterday to be unfavourable for one user might be solved differently to give that user the benefit some other time
AI and the greening of aviation		
Opportunity	**Challenge**	**Issue**
Generate greener trajectories (1)	Weather patterns are still unpredictable	
Generate greener trajectories (2)	This is a large computational problem	
Renewable energy	Metrics to determine the energy consumption of AI systems and the use of green energy for the systems	

6 Summary and Conclusion

6.1 Summary

The aim of this document, "Advanced Automation in Aviation", is to provide an up-to-date overview of the latest developments and research directions in the use of advanced automation in aviation. The focus is on the current state of research and the application of advanced automation techniques and AI in aviation. Current developments and future scenarios of automated systems and AI applications were examined and the associated opportunities, challenges, and requirements were described.

First, the document discusses the level of automation taxonomies, which is important for the categorisation of automated systems. Taxonomies from SESAR and EASA, which are relevant in aviation, are described and a recommendation for the standardisation of this taxonomy for different institutions is proposed. The research criteria for the following literature review are compiled. In the next step, the various AI methods available are discussed and the different categorisations are explained. This will also illustrate the range and diversity of AI methods that could be applied in aviation.

Next is a comprehensive literature review of automation advances for various modes of transportation, including air, rail, road, and maritime. After a discussion of general trends in mobility, specific technical trends in each mode of transportation are discussed in detail. One focus is on support systems in the areas of ATM and ATC. In summary, the goals of safety, holism, transversality, human-centeredness and human well-being are at the forefront of automation.

Advanced automation and the use of AI in aviation focus on the two main topics of airspace optimization and enhanced human support in conjunction with higher automation. A comprehensive literature review of the current trends and advances in ATM and ATC automation, directly related to the SESAR flagship "Capacity-on-Demand and Dynamic Airspace" and the use cases defined in the HUCAN project, is provided based on 13 selected technical articles. The focus here is on dynamic airspace configuration, human-autonomy teaming and the development of new decision support systems.

The document presents a detailed literature survey on current developments in the field of human assistants and some elements of current research on key aspects of these support systems. These include human-machine interaction, again human-AI teaming, trust, explainability, task assignment, and performance metrics. Human assistants thus represent a cross-cutting topic for all areas for which a theoretical framework for taxonomy, research, classification, and benchmarking is required. The functioning of the human assistant system loaded with high automation requires human trust in the system. A detailed discussion covers various traits of establishing trust in the automation systems. Additional emphasis was put on the effect of AI systems and their ability to generate human-understandable explanations (overview of the field of explainable AI and how it could be established).

Another important objective in this work package was to analyse the impact of automated systems and AI on aviation. The project focused on identifying the key opportunities and challenges, with a particular focus on the human factor and safety issues. To achieve this goal, a workshop was organised to collect and discuss "opportunities and challenges".

EUROPEAN PARTNERSHIP
Co-funded by the European Union

In the workshop, ten explicit opportunities and five challenges were identified by the participants in connection with the smart automation of functions and processes in aviation. The opportunities include, for example, more efficient data processing, systems that learn during operations and can adapt to the behaviour of air traffic controllers, thus enabling improved problem-solving behaviour and even crisis management in the long term, and the establishment of dynamic airspace adaptation for different air traffic carriers.

The challenges identified include weather phenomena, for example, which pose particular challenges for self-learning systems due to their high meteorological and traffic-related parameter variability. Efficient data processing also poses a challenge, as high data quality is required. If a support system is personalised, a fair balance between automation and human involvement must be ensured in addition to legal aspects. The use of AI represents a new way of recognising and combating cyber attacks. However, AI also offers new opportunities to carry out these attacks and thus jeopardise air traffic safety.

Automated systems that can analyse complex situations, learn and thus make decisions in new situations will change aviation forever in the near future. These systems show amazing performance in cognitive tasks and, through their integration, promise a higher level of automation in both air traffic management (ATM) and air traffic control (ATC) in order to achieve a higher level of safety, efficiency and reliability. This document provides an overview of ideas and current research on automation in aviation. Despite the opportunities presented by automated continuous learning and adaptation, challenges remain on both the technical and human side that need to be addressed by all stakeholders in order to successfully establish advanced automation.

6.2 Conclusion

One of the main goals of this WP is to collect opportunities and challenges regarding the application of high automation and AI in ATM systems. Another main gap was to suggest a unified LOAT taxonomy. The following are major conclusions from this document:

1. HUCAN high automation approach targets two main aspects of Human-AI teaming (HAT), cooperation with a directive interaction and collaboration with a focus on joint problem-solving and shared awareness.
2. To support human decision-making, information must be provided in a meaningful way to support data acquisition to decision making. To make AI decisions understandable to humans, emphasis should be put on the explainability of AI decisions.
3. Furthermore, automation should be able to identify possible errors (even human errors) to adapt to evolving situations.
4. Adaptability to high automation solutions to new environments and systems is essential for viable and safe solutions.
5. There is a clear opportunity for AI power to model complex problems with the ability to learn from multi-objective targets. However, special care should be taken in defining learning objectives for the AI algorithms, to have a holistic coverage of the problem.

References

Abduljabbar, R., Dia, H., Liyanage, S., & Asadi Bagloee, S. (2019). Applications of Artificial Intelligence in Transport: An Overview. Sustainability, 11(189), 1-24.

Ahmed, M., Mirza, M. A., Raza, S., Ahmad, H., Xu, F., Khan, W. U., Han, Z. (2023). Vehicular Communication Network Enabled CAV Data Offloading: A Review. IEEE Transactions on Intelligent Transportation Systems.

Ahrenhold, N., Gerdes, I., Mühlhausen, T., & Temme, A. (2023). Validating Dynamic Sectorization for Air Traffic Control Due to Climate Sensitive Areas: Designing Effective Air Traffic Control Strategies. Aerospace, 10(5), 405.

Ali, A., Tilbury, D. M., & Robert Jr, L. P. (2022). Considerations for Task Allocation in Human-Robot Teams. arXiv preprint arXiv:2210.03259.

Bäck, Thomas, and Hans-Paul Schwefel. (1993). "An overview of evolutionary algorithms for parameter optimization." Evolutionary computation 1.1 (1993): 1-23.

Bansal, Himanshu & Khan, Rizwan. (2018). A Review Paper on Human Computer Interaction. International Journal of Advanced Research in Computer Science and Software Engineering. 8. 53. 10.23956/ijarcsse.v8i4.630

Baribieri, G., Espana, A., & Sanchez-Londono, D. (2022). A Taxonomy for Levels of Automation based on the Industrial Revolutions. IFAC-PapersOnLine, 368-373.

Başpınar, B., Balakrishnan, H., & Koyuncu, E. (2020). Optimization-based autonomous air traffic control for airspace capacity improvement. IEEE Transactions on Aerospace and Electronic Systems, 56(6), 4814-4830.

Bazzan, Ana L.C., et al. (2014) "Symbolic Artificial Intelligence Techniques for Flight Planning." Journal of Aerospace Technology and Management, vol. 6, no. 1, 2014, pp. 93-104

Brodersen, M., Pink, S., & Fors, V. (2023). Automating the first and last mile? Reframing the 'challenges' of everyday mobilities. Mobilities.

Buchanan, B. G., & Shortliffe, E. H. (1984). Rule-based expert systems: The MYCIN experiments of the Stanford Heuristic Programming Project. Addison-Wesley

CANSO. "Operational Concept for Managing Arrival Traffic." Civil Air Navigation Services Organisation, 2019

Chiou, E. K., & Lee, J. D. (2023). Trusting Automation: Designing for Responsivity and Resilience. Human Factors.

Chromik, M., et al., 2021 - Chromik, M., & Butz, A. (2021). Human-XAI interaction: a review and design principles for explanation user interfaces. In Human-Computer Interaction–INTERACT 2021: 18th IFIP TC 13 International Conference, Bari, Italy, August 30–September 3, 2021, Proceedings, Part II 18 (pp. 619-640). Springer International Publishing.

Collins, A. and A. Evans (1994). "Aircraft Noise and Residential Property Values: An Artificial Neural Network Approach." Journal of Transport Economics and Policy 28(2): pp. 175-197.

Dasgupta, Dipankar, and Zbigniew Michalewicz, (2013) eds. Evolutionary algorithms in engineering applications. Springer Science & Business Media, 2013.

de la Pena Zarzuelo, I., Freire Soeane, M., & Lopez Bermudez, B. (2020). Industry 4.0 in the port and maritime industry: A literature review. Journal of Industrial Information Integration, 20.

Degas, A., Islam, M. R., Hurter, C., Barua, S., Rahman, H., Poudel, M., ... & Arico, P. (2022). A survey on artificial intelligence (ai) and explainable ai in air traffic management: Current trends and development with future research trajectory. Applied Sciences, 12(3), 1295.

Dekhtyaruk, M. (2023). Automated System for Freight Transportation Optimization on the Transport Network. Periodica Polytechnica Transportation Engineering, 386–393.

Dekker, S.W.A. & Woods, D.D. (2002). MABA-MABA or Abracadabra? Progress on Human -Automation Co-ordination. Cognition, Technology & Work, 4, 240-244.

Delahaye, D., & Puechmorel, S. (2008). 3D airspace design by evolutionary computation. In 2008 IEEE/AIAA 27th Digital Avionics Systems Conference (pp. 3-B). IEEE.

Doshi-Velez F, Kim B (2017) Towards a rigorous science of interpretable machine learning. arXiv e-prints abs/1702.08608. URL http://adsabs.harvard. edu/abs/2017arXiv170208608D

DW, G. D. A. (2019). DARPA’s explainable artificial intelligence program. AI Mag, 40(2), 44.

Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., ... & Ranjan, R. (2023). Explainable AI (XAI): Core ideas, techniques, and solutions. ACM Computing Surveys, 55(9), 1-33.

EASA. (2020). Artificial Intelligence Roadmap 1.0. Human-centric approach to Ai in aviation. Cologne, Germany.

EASA (2022) - Roadmap, E. A. I. (2022). EASA Concept Paper: First usable guidance for Level 1 and level 2 machine learning applications.

EASA. (2023). Artificial Intelligence Roadmap 2.0. Human-centric approach to AI in aviation. Cologne, Germany.

EC Competence Centre on Foresight. (2022). Digital Transformation. Retrieved from Knowledge for Policy: https://knowledge4policy.ec.europa.eu/foresight/digital-transformation_en

EC JRC. (2019). Digital Transformation in Transport, Construction, Energy, Government and Public Administration. Luxembourg: Publications Office of the European Union.

EC JRC. (2023). Research and Innovation in Transport Safety and Resilience in Europe. Luxembourg: Publications Office of the European Union.

EC, DG Mobility and Transport. (n.d.) (2019). STRIA Smart Mobility and Services - Roadmaps 2019. Brussels, Belgium.

EC, EU Policy Lab. (2023). Everybody is looking into the future! A literature review of reports on emerging technologies and disruptive innovation. Luxembourg: Publications Office of the European Union.

EC, JRC, TRIMIS. (2020). STRIA Roadmap on Transport Infrastructure (INF). Ispra, Italy.

EC, JRC. (2019). STRIA Roadmap on Connected and Automated Transport - Road, Rail and Waterborne. Brussels, Belgium.

EC, JRC. (2021). AI Watch - AI Uptake in Smart Mobility. Luxembourg: Publications Office of the European Union.

EC. (2020). The Sustainable and Intelligent Mobility Strategy ((COM)2020/789 final). COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS Sustainable and Smart Mobility Strategy – putting European transport on track for the future. Brussels, Belgium.

EC. (2023). COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS. Greening Freight Transport (COM(2023) 440.final). Strasbourg.

EIT Urban Mobility Innovation Hub. (2021). Urban Mobility Next #3 - AI Mobility Landscape in the EU. Barcelona, Spain.

EP. (2021a). Artificial Intelligence in smart cities and urban mobility (PE 662.937). Luxembourg.

EP. (2021b). The role of Artificial Intelligence in the European Green Deal (PE 662.906). Luxembourg.

Ersal, T., Kolmanovsky, I., Masoud, N., Ozay, N., Scruggs, J., Vasudevan, R., & Orosz, G. (2020). Connected and automated road vehicles: state of the art and future challenges. Vehicles System Dynamics, 58(5), 672-704.

EUROCONTROL (2023) “ARGOS Factsheet” https://www.eurocontrol.int/publication/argos-factsheet

Eurofound. (2022). Ethics in the digital workplace. Anticipating and managing the impact of change. Luxembourg: Publications Office of the European Union.

Feng, K., Liu, D., Liu, Y., Liu, H., & Song, H. (2023). GraphDAC: A Graph-Analytic Approach to Dynamic Airspace Configuration. In 2023 IEEE 24th International Conference on Information Reuse and Integration for Data Science (IRI) (pp. 235-241). IEEE.

Ferreira JJ, Monteiro MS (2020) What Are People Doing About XAI User Experience? A Survey on AI Explainability Research and Practice. In: Design, User Experience, and Usability. Design for Contemporary Interactive Environments. Springer International Publishing, Lecture Notes in Computer Science, pp 56{73,

Fitts,(1951). Human engineering for an effective air-navigation and traffic-control system. Washington, DC: National Research Council

Flener, P., & Pearson, J. (2013). Automatic airspace sectorization: A survey. arXiv preprint arXiv:1311.0653.

Fortunati, L., & Edwards, A. (2022). Framing the psycho-social and cultural aspects of human-machine communication. Human-Machine Communication.

Gerdes, I., Temme, A., & Schultz, M. (2018). Dynamic airspace sectorization for flight-centric operations. Transportation Research Part C: Emerging Technologies, 95, 460-480.

Giordano, D., Giobergia, F., Pastor, E., La Macchia, A., Cerquitelli, T., Baralis, E., . . . Tricarico, D. (2022). Data-driven strategies for predictive maintenance: Lesson learned from an automotive use case. Computers in Industry.

Gokasar, I., Timurogullari, A., Deveci, M., & Garg, H. (2023). SWSCAV: Real-time traffic management using connected autonomous vehicles. ISA Transactions, 24-38.

Gomez Comendador, V. F., Arnaldo Valdés, R. M., Vidosavljevic, A., Sanchez Cidoncha, M., & Zheng, S. (2019). Impact of trajectories' uncertainty in existing ATC complexity methodologies and metrics for DAC and FCA SESAR concepts. Energies, 12(8), 1559.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.

Grochow, J. M. (2020). A taxonomy of automated assistants. Communications of the ACM, 63(4), 39-41.

Gunning, D., Vorm, E., Wang, Y., & Turek, M. (2021). DARPA's explainable AI (XAI) program: A retrospective. Authorea Preprints.

Haque, A. B., Islam, A. N., & Mikalef, P. (2023). Explainable Artificial Intelligence (XAI) from a user perspective: A synthesis of prior literature and problematizing avenues for future research. Technological Forecasting and Social Change, 186, 122120.

Hedström, A., Weber, L., Krakowczyk, D., Bareeva, D., Motzkus, F., Samek, W., & Höhne, M. M. C. (2023). Quantus: An explainable ai toolkit for responsible evaluation of neural network explanations and beyond. Journal of Machine Learning Research, 24(34), 1-11.

Hernandez, C. S., Ayo, S., & Panagiotakopoulos, D. (2021). An explainable artificial intelligence (xAI) framework for improving trust in automated ATM tools. In 2021 IEEE/AIAA 40th Digital Avionics Systems Conference (DASC) (pp. 1-10). IEEE.

Heuillet A, Couthouis F, Diaz-Rodriguez N (2021) Explainability in deep reinforcement learning. Knowledge-Based Systems 214:106,685. https://doi.org/10.1016/j.knosys.2020.106685

Hilty, L. and Aebischer B. (2015). "Ict for sustainability: An emerging research field", Advances in Intelligent Systems and Computing, Vol. 310, https://doi.org/10.1007/978-3-319-09228-7_1

Holzinger, A., Goebel, R., Fong, R., Moon, T., Müller, K. R., & Samek, W. (2020, July). xxAI-beyond explainable artificial intelligence. In International Workshop on Extending Explainable AI Beyond Deep Models and Classifiers (pp. 3-10). Cham: Springer International Publishing.

Holzinger, A., Saranti, A., Molnar, C., Biecek, P., & Samek, W. (2020, July). Explainable AI methods-a brief overview. In International Workshop on Extending Explainable AI Beyond Deep Models and Classifiers (pp. 13-38). Cham: Springer International Publishing.

Hu, X. B. (2020). A methodological framework of human-machine Co-evolutionary intelligence for decision-making Support of ATM. In 2020 Integrated Communications Navigation and Surveillance Conference (ICNS) (pp. 5C3-1). IEEE.

IBM. (2021). The basics of business automation. Retrieved from https://www.ibm.com/blog/basics-of-business-automation/

Iclodean, C., Varga, B. O., & Pfister, F. (2023). Autonomous Vehicles Technological Trends. In C. Iclodean, B. O. Varga, & F. Pfister, Autonomous Vehicles Technological Trends (pp. 1-8). Basel: MDPI.

JRC. (2020). Glossary of human-centric artificial. Luxembourg, Luxembourg: Publications Office of the European Union.

Kim, J., Justin, C., Mavris, D., & Briceno, S. (2022). Data-driven approach using machine learning for real-time flight path optimization. Journal of Aerospace Information Systems, 19(1), 3-21.

Koukaki, T., & Tei, A. (2020). Innovation and maritime transport: A systematic review. 8(3), 700-710.

Krupitzer, C., Müller, S., Lesch, V., Züfle, M., Edinger, J., Lemken, A., ... & Becker, C. (2020). A survey on human machine interaction in industry 4.0. arXiv preprint arXiv:2002.01025.

Kulida, E., & Lebedev, V. (2020). About the Use of Artificial Intelligence Methods in Aviation. IEEE.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.

Lim, W. M. (2023). The workforce revolution: Reimagining work, workers, and workplaces for the future. Global Business and Organisational Excellence.

López, M. M., & Gutiérrez, M. G. (2023). Collaborative ATM-U-Space Interface (No. 11032). EasyChair.

Ly, H.-B. et al. (2019) Development of an AI Model to Measure Traffic Air Pollution from Multisensor and Weather Data. Sensors (2019), 19(22):4941. https://doi.org/10.3390/s19224941

Lyons, J. B., K. Sycara, M. Lewis and A. Capiola (2021). "Human–Autonomy Teaming: Definitions, Debates, and Directions." Frontiers in Psychology 12.

Malakis, S., M. Baumgartner, N. Berzina, T. Larsen, A. Smoker, A. Poti and G. Fabris (2022). "Challenges from the Introduction of Artificial Intelligence in the European Air Traffic Management System." IFAC-PapersOnLine 55(29): pp. 1-6.

Movila, G. D. (2023). Scoring mechanism for automated ATC systems (Bachelor's thesis, Universitat Politècnica de Catalunya).

Mueller ST, Veinott ES, Homan RR, et al (2021) Principles of explanation in human-AI systems. CoRR abs/2102.04972. URL https://arxiv.org/abs/2102.04972

Murphy, K. P. (2012). Machine learning: a probabilistic perspective. MIT press.

Naikar, N. (2018). Human–automation interaction in self-organizing sociotechnical systems. Journal of Cognitive Engineering and Decision Making, 12(1), 62–66. doi:10.1177/1555343417731223

Naikar, N., & Elix, B. (2016). Integrated system design: Promoting the capacity of sociotechnical systems for adaptation through extensions of cognitive work analysis. Frontiers in Psychology, 7, 1–21. doi:10.3389/fpsyg.2016.00962

Naikar, N., Elix, B., Dâgge, C., & Caldwell, T. (2017). Designing for self-organisation with the diagram of work organization possibilities. In J. Gore & P. Ward (Eds.), Proceedings of the 13th international conference on Naturalistic Decision Making (pp. 159–166). Bath, UK: University of Bath.

OECD (2022), "Measuring the environmental impacts of artificial intelligence computing and applications: The AI footprint", OECD Digital Economy Papers, No. 341, OECD Publishing, Paris, https://doi.org/10.1787/7babf571-en.

Parasuraman, R., Sheridan, T.B., & Wickens, C.D. (2000). A model for types and levels of human interaction with automation. IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans, 30, 286–297.

Phillips, P. J., Hahn, C. A., Fontana, P. C., Broniatowski, D. A., & Przybocki, M. A. (2020). Four Principles of Explainable Artificial. NIST Interagency/Internal Report (NISTIR). Gaithersburg, MD: National Institute of Standards and Technology.

Puiutta E, Veith EMSP (2020) Explainable reinforcement learning: A survey. In: Machine Learning and Knowledge Extraction - 4th IFIP TC 5, TC 12, WG 8.4, WG 8.9, WG 12.9 International Cross-Domain Conference, CD-MAKE 2020, Dublin, Ireland, August 25-28, 2020, Proceedings, Lecture Notes in Computer Science, vol 12279. Springer, pp 77{95, https: //doi.org/10.1007/978-3-030-57321-8 5

Ravuri, S. et al. (2021), “Skilful precipitation nowcasting using deep generative models of radar”, Nature 597, pp. 672–677, https://doi.org/10.1038/s41586-021-03854-z.

Ren, F., & Bao, Y. (2020). A Review on Human-Computer Interaction and Intelligent Robots. Int. J. Inf. Technol. Decis. Mak., 19, 5-47

Roth, E. M., Sushereba, C., Militello, L. G., Diiulio, J., & Ernst, K. (2019). Function allocation considerations in the era of human autonomy teaming. Journal of Cognitive Engineering and Decision Making, 13(4), 199-220.

Sachendra Yadav, (2023) “Function allocation can optimize performance, efficiency, safety, and user satisfaction” - https://bootcamp.uxdesign.cc/function-allocation-is-the-key-to-human-ai-teaming-in-enterprise-workflows-4fcdf150e455

Saeed, W., & Omlin, C. (2023). Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities. Knowledge-Based Systems, 263, 110273.

Samad, T. (2023). Human-in-the-Loop Control and Cyber–Physical–Human Systems: Applications and Categorization. In A. M. Annaswamy, P. P. Khargonekar, F. Lamnabhi-Lagarrigue, & S. K. Spurgeon, Cyber–Physical–Human Systems: Fundamentals and Applications. Hoboken, New Jersey: John Wiley & Sons, Inc.

Sánchez, J. A. (1997). A taxonomy of agents. Rapport technique, ICT-Universidad de las Américas-Puebla, México

Save, L. & Feuerberg, B. (2012). Designing Human-Automation Interaction: a new level of Automation Taxonomy, in D. de Waard, K. Brookhuis, F. Dehais, C. Weikert, S. Röttger, D. Manzey, S. Biede, F. Reuzeau, and P. Terrier (Eds.), Human Factors: a view from an integrative perspective. Proceedings HFES Europe Chapter Conference Toulouse. ISBN 978-0-945289-44-9. Available from http://hfes-europe.org.

Schwalbe, G., & Finzel, B. (2023). A comprehensive taxonomy for explainable artificial intelligence: a systematic survey of surveys on methods and concepts. Data Mining and Knowledge Discovery, 1-59.

SEI, (2022). Perceptions of Function Allocation between H and AI-E Systems

SESAR (2013). SESAR Joint Undertaking, 'Guidance Material for HP Automation Support', 16.5.1, D04, 2013 (https://ec.europa.eu/research/participants/data/ref/h2020/other/guides_for_applicants/h2020-sesar-er4-26-2019_d04_condensed_16.05.01_en.pdf)

SESAR. (2020). Automation in air traffic management. Long-term vision and initial research roadmap. Luxemburg, Luxemburg: Publications Office of the European Union.

Sheridan, T.B., & Verplank, W. (1978). Human and Computer Control of Undersea Teleoperators. Cambridge, MA: Man-Machine Systems Laboratory, Department of Mechanical Engineering, MIT.

Speith, T. (2022). A review of taxonomies of explainable artificial intelligence (XAI) methods. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (pp. 2239-2250).

Stefan van Duin, N. B. (2018). Artificial intelligence whitepaper. Technical report, Deloitte The Netherlands

Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction. MIT press.

Swamy, V., Frej, J., & Käser, T. (2023). The future of human-centric eXplainable Artificial Intelligence (XAI) is not post-hoc explanations. arXiv preprint arXiv:2307.00364.

Tang, H., Zhang, Y., Mohmoodian, V., & Charkhgard, H. (2021). Automated flight planning of high-density urban air mobility. Transportation Research Part C: Emerging Technologies, 131, 103324.

Tang, J., G. Liu and Q. Pan (2022). "Review on artificial intelligence techniques for improving representative air traffic management capability." Journal of Systems Engineering and Electronics 33(5): pp. 1123–1113.

Tang, R., De Donato, L., Besinovic, N., Flammini, F., Goverde, R. M., Lin, Z., Wang, Z. (2022). A literature review of Artificial Intelligence applications in railway systems. Transport Research Part C: Emerging Technologies, 140.

Teutsch, J., Petersen, C., Schwoch, G., Lieb, T. J., Bos, T., & Zon, R. (2023). On the Impact of UAS Contingencies on ATC Operations in Shared Airspace. In 2023 Integrated Communication, Navigation and Surveillance Conference (ICNS) (pp. 1-15). IEEE.

The Royal Society, (2019) - Explainable AI: the basics Policy briefing Issued: November 2019 DES6051 ISBN: 978-1-78252-433-5

Vouros, 2022 - Vouros, G. A. (2022). Explainable deep reinforcement learning: state of the art and challenges. ACM Computing Surveys, 55(5), 1-39.

Xiong, W., Fan, H., Ma, L., & Wang, C. (2022). Challenges of human—machine collaboration in risky decision-making. Frontiers of Engineering Management, 9(1), 89-103.

Xu, C., Zhang, K., Liu, D., & Song, H. (2021). Trust of Airspace Configuration Transition Concerning the Fluctuation of Air Traffic. In 2021 IEEE/AIAA 40th Digital Avionics Systems Conference (DASC) (pp. 1-7). IEEE.

Yang et al., 2022 Yang, J., Liu, T., Liu, Y., & Morgan, P. (2022). Review of human-machine interaction towards industry 5.0: human-centric smart manufacturing. In International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (Vol. 86212, p. V002T02A060). American Society of Mechanical Engineers

Zhou J, et al., 2021 - Zhou J, Gandomi AH, Chen F, et al (2021) Evaluating the quality of machine learning explanations: A survey on methods and metrics. Electronics 10(5):593. https://doi.org/10.3390/electronics10050593

Zhou, W., Cai, Q., & Alam, S. (2022). A multi-task learning approach for facilitating dynamic airspace sectorization. In Proceedings of International Workshop on ATM/CNS 2022 International Workshop on ATM/CNS (pp. 192-199). Electronic Navigation Research Institute.

List of acronyms

Table 9 List of acronyms

Acronym	Description
AA	Advanced Automation
ATC	Air Traffic Control
AI	Artificial Intelligence
AMAN	Arrival Manager
ANSP	Air Navigation Service Provider
ARGOS	ATC Real Ground-breaking Operational System
ATCO	Air Traffic Controllers
ATM	Air Traffic Management
CTA	Cognitive Task Analysis
CWA	Cognitive Work Analysis
DA	Digital Assistant
DMAN	Departure Manager
EASA	European Union Aviation Safety Agency
EEA	European Economic Area
eVTOL	electric vertical take-off and landing
EU	European Union
HAT	Human Autonomy Teaming
HITL	Human-in-the-Loop
HMI	Human Machine Interface
ICAO	International Civil Aviation Organization
KPI	Key Performance Indicator
LOAT	level of automation taxonomy
ML	Machine Learning
KPI	key performance indicator
SESAR	Single European Sky ATM Research Programme
SJU	SESAR Joint Undertaking (Agency of the European Commission)

SRIA	Strategic Research and Innovation Agenda
TMA	Terminal Manoeuvring Area
UAM	Urban Air Mobility
UAV	Unmanned Aerial Vehicles
WP	Work Package
XAI	Explainable AI