Technische Universität München

Institute of Flight System Dynamics

Towards a Regulatory Framework for the Certification of AI/ML-based Airborne Systems in Military Aviation

–

Auf dem Weg zu einem Regelungsrahmen für die Zertifizierung von KI-basierten Luftfahrtsystemen in der militärischen Luftfahrt

Masterarbeit

Author: Clara Maria Capdevila Llompart

Matriculation Number: 03770351

Supervisor: Prof. Dr.-Ing. Florian Holzapfel

M. Sc. Konstantin Dmitriev

Dr.-Ing. Antonio Monzón Díaz

November 2023

Statutory Declaration

I, Clara Maria Capdevila Llompart, declare on oath towards the Institute of Flight System Dynamics of Technische Universität München, that I have prepared the present Master thesis independently and with the aid of nothing but the resources listed in the bibliography.

This thesis has neither as-is nor similarly been submitted to any other university.

Garching, 02 November 2023

Clara Maria Capdevila Llompart

Kurzfassung

Künstliche Intelligenz (KI), insbesondere maschinelles Lernen (ML), gewinnt zunehmend an Bedeutung im Luft- und Raumfahrtsektor, da sie zahlreiche Vorteile bietet und Innovationen vorantreibt. Die einzigartigen Eigenschaften von AI/ML-basierten Luftfahrtsystemen stellen jedoch Herausforderungen bei der Integration in herkömmliche Design- und Zertifizierungsstandards dar. Dies hat dazu geführt, dass im zivilen Bereich Bemühungen unternommen wurden, bestehende Vorschriften zu aktualisieren. Im militärischen Luftfahrtbereich hinken diese Bemühungen jedoch hinterher. Militärische Luftfahrt stellt andere Herausforderungen als die zivile Luftfahrt dar, da sie sicherheitskritische Funktionen mit spezifischen Leistungsanforderungen umfasst, die sich aus unterschiedlichen Missionszielen, Bedrohungsumgebungen und Zertifizierungsrahmen ergeben. AI/ML-basierte militärische Avioniksysteme müssen Faktoren wie Luftkampfmissionen, Ausweichmanöver, Zielidentifikation und die Integration von Waffensystemen berücksichtigen. Diese Studie konzentriert sich auf die Analyse des regulatorischen Rahmens für AI/ML-basierte militärische Systeme. Die Hauptergebnisse dieser Arbeit umfassen eine Überprüfung der Bemühungen zur zivilen Standardisierung, eine Kategorisierung militärischer AI/ML-Systeme, eine Analyse militärspezifischer Zertifizierungsherausforderungen, eine Bewertung der Anwendung ziviler Bemühungen auf militärische Kontexte und eine Berücksichtigung der Entwicklung militärspezifischer Zertifizierungsziele. Diese Forschung zielt darauf ab, Bereiche zu identifizieren, in denen die militärische Luftfahrt möglicherweise einen anderen Ansatz zur Zertifizierung von AI/ML-Systemen erfordert, wobei der Schwerpunkt auf Sicherheit, Zuverlässigkeit, operationellem Kontext, Mensch-Maschine-Interaktion, regulatorischer Compliance und ethischen Überlegungen liegt.

Abstract

Artificial Intelligence (AI), in particular Machine Learning (ML), is becoming increasingly important in the aerospace sector, offering numerous benefits and driving innovation. However, the unique characteristics of AI/ML-based airborne systems pose challenges in integrating them into traditional design assurance standards and certification approaches. This has led to efforts in the civil world to update existing regulations. However, these efforts have lagged behind in military aviation. Military aviation presents different challenges from civil aviation, as it involves safety-critical functions with specific performance requirements derived from different mission objectives, threat environments and certification frameworks. AI/ML-based military avionics systems must consider factors such as air combat mission collaboration, evasive manoeuvres, target identification and weapon system integration. This study focuses on the analysis of the regulatory framework for AI/ML-based military systems. The main outputs of this work include a review of civil standardisation efforts, a categorisation of military AI/ML systems, an analysis of military-specific certification challenges, an assessment of the application of civil efforts to military contexts, and a consideration of the development of military-specific certification targets. This research aims to identify areas where military aviation may require a different approach to the certification of AI/ML systems, focusing on safety, reliability, operational context, human-machine interaction, regulatory compliance, and ethical considerations.

Table of Contents

List of Figures

List of Tables

Table of Acronyms

Acronym	Description
A3R	Automatic Air-to-Air Refuelling
AAR	Air-to-Air Refuelling
AI	Artificial Intelligence
AI4DEF	AI for Defence
ACAS	Airborne Collision Avoidance System
AMC	Acceptable Means of Compliance
ALTAI	Assessment List for Trustworthy AI
ARBS	Aerial Refuelling Boom System
ARO	Air Refuelling Officer
ATM	Air Traffic Management
AIA	Artificial Intelligence Act
C2	Command and Control
C4I	Command, Control, Computers, Communications, and Intelligence
CAA	Civil Airworthiness Authorities
CNN	Convolutional Neuronal Network
CoD	Certificate of Design
ConOps	Concept of Operations
CQC	Certification and Qualification Committee
CQO	Certification and Qualification Organisation
CRI	Certification Review Item
CS	Certification Specification
CSR	Cyber Security and Resilience
DAAI	Defence Applications of AI
DAL	Design Assurance Level
DARB	Data and Artificial Intelligence Review Board

DARPA	Defence Advanced Research Projects Agency
DASS	Defence Aids Systems
DGAC	France is the Direction Générale de l'Aviation Civile
DL	Deep Learning
DOA	Design Organization Approval
DoD	U.S. Department of Defence
EASA	European Aviation Safety Agency
EDA	European Defence Agency
EICACS	European Initiative Collaborative Air Combat Standardization
ESM	Electronic Support Measures
EU	European Union
EUROCAE	European Organization for Civil Aviation Equipment
EW	Electronic Warfare
FAA	Federal Aviation Administration
FCAS	Future Combat Air System
FMS	Flight Management System
FSD	Flight System Dynamics Institute
F2T2EA	Find, Fix, Track, Target, Engage and Asses
HAT	Human-AI Teaming
HMD	Head-Mounted Display
HUD	Head-up Display
IPC	Innovation Partnership Contract
ISO	International Standards Organisation
ISR	Intelligence, Surveillance, and Reconnaissance
KIEZ	Künstliche Intelligenz Europäisch Zertifizieren
LAWS	Lethal autonomous Weapons Systems
LBA	Luftfahrt-Bundesamt
LufABw	Luftfahrtamt der Bundeswehr

MAA	Military Airworthiness Authority
MALE	Medium Altitude Long Endurance
MCRI	Military Certification Review Item
ML	Machine Learning
MLEAP	Machine Learning Application Approval
MOC	Methods of Compliance
MoC	Memoranda of Cooperation
MPA	Maritime Patrol Aircraft
MPE	Mission Planning and Execution
MRCA	Multirole Combat Aircraft
MRO	Maintenance Repair and Overhaul
NATO	North Atlantic Treaty Organization
NIAG	NATO Industrial Advisory Group
NN	Neuronal Network
OCCAR	Organisation for Joint Armament Cooperation
OODA	Observe, Orient, Decide and Act
PRU	Principles of Responsible Use
RPAS	Remote Pilot Aircraft Systems
SARP	Standards and Recommended Practises
SAE	International Society of Automotive Engineers
SDLC	Software Development Life Cycle
STANAG	Standardisation Agreement (NATO)
STC	Science and Technology Committee (NATO)
STO	Science and Technology Organization (NATO)
VLS	Visual Landing System
UAV	Unmanned Aerial Vehicle

Table of Symbols

Latin Letters

Symbol	Unit	Description
d	m	Distance to the receiver's intake
h	–	Hypothesis function
S	–	Dataset
U	–	Decision space
u	–	Output value
X	–	Input space
x	–	Input value

Greek Letters

Symbol	Unit	Description
β	°	Boom pitch
θ	–	Parameters of the hypothesis function
φ	°	Boom yaw

Indices

Symbol	**Description**
i	index

1 Introduction

1.1 Background

Over the past few decades, artificial intelligence (AI) has become an indispensable and transformative technology that has reshaped many industries and promises to revolutionise even more, including aviation. The expected implementation of AI into avionics holds great promise for modifying aircraft operation, ensuring safety, and optimising performance. From autonomous flight systems and advanced navigation to real-time data analysis for in-flight decision making, the presence of AI in avionics is elevating aviation to new heights. In stark contrast to the standardised and often routine missions of civil aviation, military aviation is an area where AI is not just a technological advance, but an innovative force that pushes the boundaries of what is possible. The defence sector's use of AI is characterised by its exceptional complexity and the need to perform extraordinary tasks in challenging environments. In this context, AI-based systems in military aviation have the potential to profoundly transform military operations and advanced capabilities. This potential becomes apparent as more autonomous vehicles demonstrate their ability to operate in dangerous and difficult-to-access areas, and as intelligent defensive systems demonstrate their effectiveness in detecting, analysing, and responding quickly to attacks, outperforming human operators in both speed and efficiency. AI-based systems can sift through large volumes of data and quickly detect patterns and anomalies that might escape human analysts. This capability is essential in identifying potential threats, enabling interoperability, and improving overall situational awareness. It allows military decision makers to make more informed and faster decisions, ultimately leading to better courses of action. In flight formation for example, AI algorithms enable aircraft to coordinate and manoeuvre seamlessly, executing complex aerial formations that are not only visually impressive, but also tactically advantageous. These military applications are proof to the precision and adaptability that AI brings to the skies [1].

1.2 Research problem

Over time, AI algorithms have become increasingly sophisticated, enabling improved situational awareness, better response times and reduced pilot workload. However, they also introduce complex challenges related to system reliability, accountability, and ethical considerations. As aviation, both civil and military, embrace AI-driven innovations, it becomes imperative to establish a comprehensive regulatory framework to ensure their safe, reliable, and ethical integration.

In traditional airborne systems, software development and certification for safety critical functions have followed well-established assurance approaches. Yet, the integration of AI algorithms introduces complexities that challenge the conventional certification process. For example, machine learning (ML) algorithms, which are a subset of AI, exhibit a unique characteristic when compared to traditional software. Their behaviour is data-driven, making it challenging to establish a clear and direct two-way traceability between software requirements and the software code. ML algorithms rely heavily on large datasets for training, and any changes in the data can lead to variations in the algorithm's outputs. This complex dependency relationship between requirements, data, ML models, and the implementing software calls for a different approach to verify and assess the safety of ML-based systems [2]. Consequently, the traditional requirements-based verification alone is no longer sufficient to ensure the correct and safe behaviour of AI/ML systems in safety critical domains. To address this challenge, various standardisation efforts have been initiated in the civil world to update and enhance existing regulatory frameworks for AI/ML-based systems, aiming to ensure the safety and reliability of these advanced systems in critical contexts [3].

The development, implementation, and certification of AI/ML-based military systems in safety critical contexts, presents unique challenges and requires careful consideration due to specific operational requirements, such as specific mission objectives, combat scenarios, evasive manoeuvres, target identification, and seamless defensive systems integration. Therefore, the development of a tailored certification framework for AI/ML applications in military aviation is crucial to ensure safety and compliance in this critical domain. However, it is noteworthy that very few efforts have been directed toward addressing these challenges in the realm of military avionics.

This work aims to address the following objectives:

- **Knowledge Compilation:** Compile a comprehensive literature review encompassing ongoing efforts to standardise AI/ML technology in military aviation, with a particular focus on airborne applications.
- **Identification of AI applications in military aviation:** Identify and categorise various applications of AI in military avionics systems, shedding light on their potential impact and significance.
- **Certification Challenges Assessment:** Identify and analyse the unique challenges associated with the certification of systems employing AI/ML in military aviation, distinct from the common challenges encountered in civil aviation.
- **Case Study Integration:** Contribute to the research effort by providing a case study involving a military system that integrates an AI/ML component, offering practical insights and considerations.
- **Standardisation Alignment:** Evaluate the applicability of conventional and emerging AI/ML civilian standards in the military aviation sector, with a focus on harmonising standards between both military and civilian domains wherever feasible.
- **Military-Specific Objectives:** Assess whether there is a need to develop new military-specific objectives in addition to those established in the civil domain. Recommendation of guidance material specific to military AI/ML-based avionics when the alignment is not feasible anymore.

The following objectives have been reformulated into research questions that are expected to be addressed and answered by the end of the study:

- **RQ1.** What are the ongoing efforts and initiatives aimed at standardising AI/ML technology in civil and military aviation, particularly in avionics?
- **RQ2.** What are the various applications of AI/ML in military avionics, and how do they impact and contribute to the aviation domain?
- **RQ3.** What are the specific challenges associated with certifying systems that incorporate AI/ML in military aviation, and how do these challenges differ from those encountered in civil aviation?

- **RQ4.** How can a case study involving a military system that integrates an AI/ML component contribute to the broader research effort, offering practical insights and considerations for real-world implementation?
- **RQ5.** To what extent can conventional and emerging civilian standards be applied or adapted to the military aviation sector, with a focus on harmonising standards between military and civilian domains where feasible?
- **RQ6.** Is it necessary to develop new objectives in the military domain to complement the objectives established in the civil domain? If so, what would be the specific areas where these objectives should be focused to address the unique considerations and requirements of military aviation?

To achieve this, the research methodology involves leveraging interactions and engagement with experts in the field of airworthiness in the military sector, analysing relevant publications and policy papers from reputable sources such [4]–[8], and outlining a framework for military certification of AI-based systems. Additionally, insights from both commercial and military AI/ML solution providers are incorporated to gain comprehensive perspectives on applications in military aviation. The research conducts a thorough review of relevant literature and existing regulatory materials to identify gaps or areas where the certification of AI/ML systems in military aviation may require unique considerations compared to civil aviation. Furthermore, specific objectives for AI/ML application in military aviation will be proposed, focusing on safety and reliability, operational context, human-machine interaction, regulatory compliance, and ethical considerations. By exploring the unique characteristics and operational requirements of military aircraft, the study aims to contribute to the development of an effective certification framework that facilitates the safe and efficient utilisation of AI/ML systems in this specialised domain.

1.3 Assumptions and limitations

Assumptions for this research include that the study will only assess and analyse the regulatory framework of military aviation in the European Union (EU) and the United States, which are both members of the North Atlantic Treaty Organisation (NATO). The EU and the United States play a key role in the development of international aviation standards and are renowned for having some of the safest aviation industries in the world. Their importance is due to their economic relevance, regulatory expertise, collaborative efforts, technological leadership, investment in safety, transparency and accountability in regulatory processes, and the strong safety culture that permeates their aviation sectors. These two regions contribute substantially to making aviation one of the safest forms of transportation in the world, and their standards and practices often serve as models for other countries seeking to improve aviation safety. Another assumption of this research is that, although it is recognised that AI/ML-based systems have significant potential in other aviation domains, such as air traffic management (ATM) and predictive maintenance, the scope is primarily focused on avionics systems. This research assumes that the interchangeability of the terms "AI" and "ML" is valid in the context of this study. Recognising that ML is a subdomain of AI, this thesis focuses primarily on ML techniques, acknowledging the integral role they play in the broader field of AI.

One of the primary limitations of this study is the restricted access to data, primarily due to the sensitive nature of the military domain. Data limitations in a thesis about military aviation are primarily driven by concerns related to national security, confidentiality, and sensitivity. The military aviation domain frequently deals with classified and sensitive information, making it challenging to access crucial data for research purposes. Strict security protocols and clearance requirements further hinder data collection by limiting access to military facilities and personnel. Additionally, concerns about compromising military capabilities or strategies through data disclosure restrict the availability of certain information. Diplomatic considerations, international relations, and commercial interests also play a role in restricting data sharing. These factors collectively create significant barriers to obtaining comprehensive and detailed data for research in this field. Consequently, the study may face challenges in obtaining comprehensive and detailed information, particularly regarding classified projects and initiatives, which can impact the depth and scope of the research. To address this issue, this research employs a multi-faceted strategy. It combines engagement with in-house experts who are well versed in the field of military aviation to provide valuable insights and perspectives that are not publicly available, allowing for a more complete understanding of trends, challenges, and developments. In addition, a qualitative case study, based on unrestricted civil aviation research, serves as a benchmark for comparison, allowing alignment with civil aviation practices and regulations. The approach underlines the importance of transferring best practices from the civil aviation sector to the military, potentially improving safety, efficiency, and regulatory frameworks. Furthermore, to fill data gaps, the research integrates fictional data in a realistic way, based on civilian case studies, to simulate possible military scenarios. This approach ensures credibility while exploring hypothetical but practical situations that provide valuable insights into military aviation challenges and solutions.

1.4 Outline

This thesis focuses on the need for a certification framework derived from the integration of AI/ML-based military avionic systems. The Introduction chapter of this thesis provides a comprehensive background on the integration of AI in aviation, with a particular focus on its transformative impact on military aviation. The chapter also outlines the specific research problem related to the certification of AI/ML-based systems in military aviation, setting the stage for the study's objectives and research questions. Furthermore, it acknowledges the assumptions and limitations that guide the research, particularly regarding data accessibility in the sensitive military domain.

To provide a solid theoretical foundation, the Fundamentals chapter introduces AI/ML technology in aviation. It defines key terms and explores various AI techniques employed in aviation. It also delves into the levels of autonomy exhibited by AI systems and examines their various applications in the military domain. In addition, this chapter contrasts civil aviation and military aviation certification frameworks, with the objective of identifying the specific challenges posed by the certification of AI/ML systems for military use.

The State of Science section conducts an in-depth analysis of existing approaches to AI/ML technology and certification within civil and military aviation. Extensive research is conducted to review pertinent literature from reputable sources. The goal is to classify and evaluate the effectiveness of these approaches in the context of military aviation.

Moving on to the Methodology chapter, the research questions are addressed through data collection and analysis methods, with a particular focus on developing a regulatory framework workflow to certify AI/ML-based military systems.

In the Case study section, the transferability of AI civil guidance material to military platforms is conducted. A specific military system with an integrated AI/ML component is being closely examined, with the aim of providing practical insights and considerations for the application of AI in military aviation.

Finally, the Discussion chapter summarises the main findings of the research, reiterates the research question's significance, and discusses how these questions have been addressed throughout the study. The chapter highlights the contributions of the research to the field of military aviation and AI/ML certification while providing recommendations for future research.

2 Fundamentals

This chapter provides a comprehensive exploration of the fundamental aspects of AI technology in the field of military aviation and addresses the challenges associated with its certification.

It begins by introducing the concept of AI in the aviation industry and aims to clarify the key concepts and terminology associated with AI, providing a solid foundation of knowledge. It explores the diverse range of applications of AI technology within this dynamic military domain. By highlighting these applications, an idea of the tangible benefits and potential challenges associated with the implementation of AI in military aviation is gained.

Subsequently, the focus shifts to the certification framework in military aviation, comparing it with the framework in civil aviation. It thoroughly analyses the various stakeholders involved, including regulatory authorities, industry organisations, and certification bodies. Additionally, it clarifies the process framework for certification, highlights the responsibilities of the Type Certificate (TC) holder, and delves into the specific missions and operational requirements that significantly influence the certification process. It provides valuable insights into the applicable standards and guidance materials that serve as a reference for the certification of AI/ML systems in both civil and military aviation domains.

Lastly, it identifies the challenges that arise during the certification of AI systems in military aviation. Given the distinctive operational and mission requirements of military aircraft, the certification process encounters additional complexities. These challenges are examined in detail, such as ensuring robust security and resilience against adversarial attacks, seamless integration with existing legacy platforms, effective human-machine teaming, and the certification of AI/ML-based autonomous capabilities.

By thoroughly exploring these fundamental aspects, certification frameworks, and associated challenges pertaining to AI/ML technology in aviation, the objective of this chapter is to provide readers with an in-depth understanding of the complexities involved in the seamless integration of AI into military avionic systems.

2.1 Introduction to AI technology

The integration of AI technology in the military domain has revolutionised the defence sector by enabling advanced capabilities, enhancing operational efficiency, and improving situational awareness. This chapter provides an overview of the application of AI technology in the field of military aviation. It begins with an introduction to the fundamental definitions related to AI, establishing a common understanding of the terminology used in subsequent discussions. The chapter then refers to AI in the context of classes of technology, levels of autonomy, and specific applications in the military domain. These constitute the different criteria for the classification of AI types.

2.1.1 AI/ML technology

AI is a relatively old field of computer science that encompasses several techniques and covers a wide spectrum of applications. AI is a broad term and its definition has evolved as technology developed. According to [9] AI refers to *"the development of computer systems that can perform tasks that typically require human intelligence, such as perception, reasoning, learning, and decision-making"*. For the version 2.0 of AI Roadmap [10], the European Aviation Safety Agency (EASA) has moved to the even wider-spectrum definition from the "Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence" (EU AI Act) [11], that is "*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"*.

This fundamental definition can then be altered to encompass the nuances of AI as it is used in different domains. Starting from the basic definition, for the military the definition can be further refined to read as**,** "*Intelligence introduced into a warfighting system, which provides that system with the ability to function to varying levels of autonomy and achieve a desired outcome without any human inputs for the full span of an independent mission*" [12].

With respect to different AI techniques, the NATO Science and Technology Organisation (STO) [9], [6] and the U.S. Defence Advanced Research Projects Agency (DARPA) [7], agencies focused on high-demand research and development defence projects within the NATO and the U.S. Department of Defence (DoD) respectively, categorise AI innovation into several waves (see Figure 2-1).

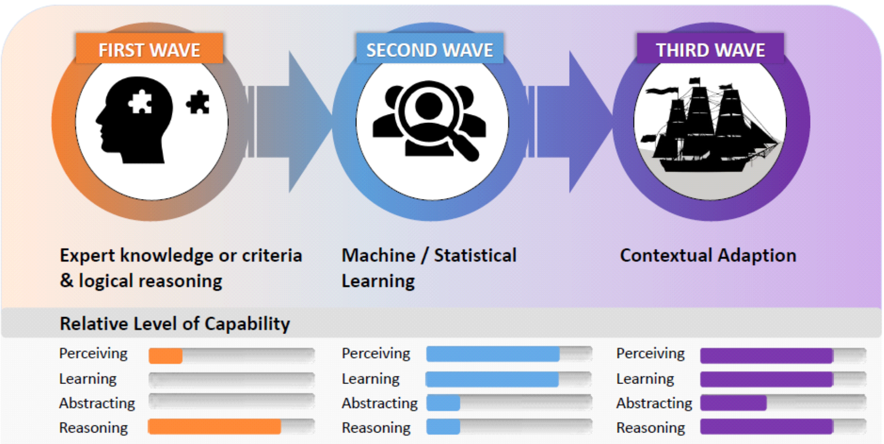

Figure 2-1: AI waves classification [8], [9].

Initially, in the 'first-wave' or 'knowledge-based' AI, the emphasis was on rule-based approaches, utilising techniques such as decision trees, Boolean and fuzzy logic. These early systems, known as expert systems, aimed to capture human expertise and knowledge [9]. First-wave systems continue to yield benefits and are an active area of AI. Expert systems are strong with respect to reasoning, as they reflect the logic and rules that are programmed into them. Expert systems are not strong, however, when it comes to perceiving, learning, or abstracting to a domain outside the one programmed into the system [8].

In the second stage, AI developers shifted their focus to statistical methods, leading to the emergence ML. ML involves training a computer system using large datasets to identify patterns, relationships, and insights, allowing it to generalise and make accurate predictions or perform tasks based on new, unseen data. This approach proved fruitful, yielding solutions like spam filters for emails and powerful internet search engines. It includes voice recognition, natural language processing, computer vision, etc. This is called ‘second wave’ or ‘data-based’ AI [9]. ‘Generalisation’ in ML refers to the ability of a trained model to perform well on new, previously unseen data that wasn't part of its training dataset. In other words, a ML model is said to generalise effectively if it can extract meaningful patterns, relationships and trends from the training data and apply them to new, similar data instances. This contrasts with traditional knowledge-based systems, where explicit rules or programmed instructions govern the system's behaviour. In these systems, adapting to new scenarios often requires manual intervention to update the rules, making them less flexible in dealing with different or evolving situations.

The future ‘third wave’, also referred to as ‘contextual adaptation’ is underway. This phase combines the strengths of ‘first-wave’ and ‘second wave’. It is characterised by AI systems that are not only capable of adapting to new situations, but also are able to abstract, contextualise and explain to users the reasoning behind these decisions [9].

On the civil side, AI is also classified regarding the different techniques’ approaches. The EU commission in its Proposal for an EU AI Act [11] and EASA in its new roadmap 2.0 [10], divide AI techniques into ML approaches (also known as data-driven AI), logic- and knowledge-based approaches (also known as symbolic AI), and statistical approaches. Consequently, the EASA first AI Roadmap [13] has been extended to encompass all techniques and approaches described in the following Figure 2-2.

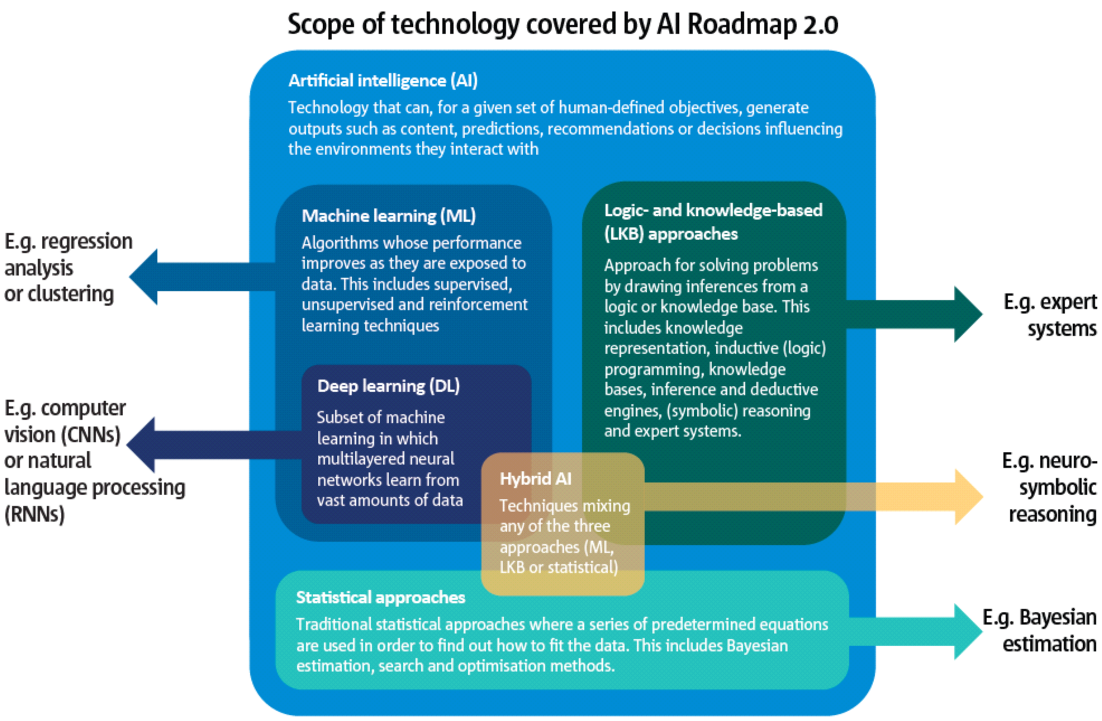

Figure 2-2: AI covered by AI Roadmap 2.0 [10].

ML approaches include supervised, unsupervised and reinforcement learning. These methods use a wide variety of techniques, most notably deep learning (DL). DL is based on neural networks (NN) with multiple layers capable of learning hierarchical representations of data [10], [11]. Next, logic and knowledge-based approaches are another important category. They include knowledge representation, inductive (logic) programming, knowledge bases, inference and deduction engines, symbolic reasoning, and expert systems. These methods focus on formalising knowledge so that machines can understand and reason with it [10], [11]. The third category encompasses statistical approaches such as Bayesian estimation and search and optimisation methods. Bayesian estimation uses probability theory to update beliefs and make predictions based on prior knowledge and observed data. Search and optimisation methods aim to find optimal solutions by exploring a solution space or approximating those using heuristics and algorithms [10], [11]. Finally, the hybrid AI approach, which involves the integration of multiple techniques and approaches to solve complex problems. It involves combining ML, logic-based reasoning, and statistical methods to leverage the strengths of hybrid AI [10], [11] In addition to the combination of different AI techniques, the term 'hybrid AI' can also refer to an approach that harmoniously blends AI methods with established, traditional techniques. This fusion of approaches aims to leverage the respective strengths of AI and traditional algorithms to create synergistic solutions to complex problems.

When drafting the first version of EASA AI Roadmap [13], the focus was primarily on ML and, specifically, DL. While learning solutions continue to dominate in the applications and use cases received from the aviation industry, it has become apparent that meeting the high safety standards set by current aviation regulations is leading certain applicants to adopt a renewed set of knowledge-based AI. Moreover, it is important to note that these different AI approaches may be used in combination, as hybrid AI.

Nevertheless, the main efforts yet published are still focused on developing preliminary guidance for implementing ML and DL components in avionics, either to introduce expert systems or statistical approaches. That is why it is meaningful to take a deeper look into ML characteristics.

<u>Machine learning (ML): data-driven AI</u>

ML is a subset of AI that involves teaching computers to learn from data and make predictions. ML algorithms identify patterns in data to create models for decision-making. It encompasses a wide range of applications, methods, and algorithms. Figure 2-3 schematises the main differences between ML and traditional programming.

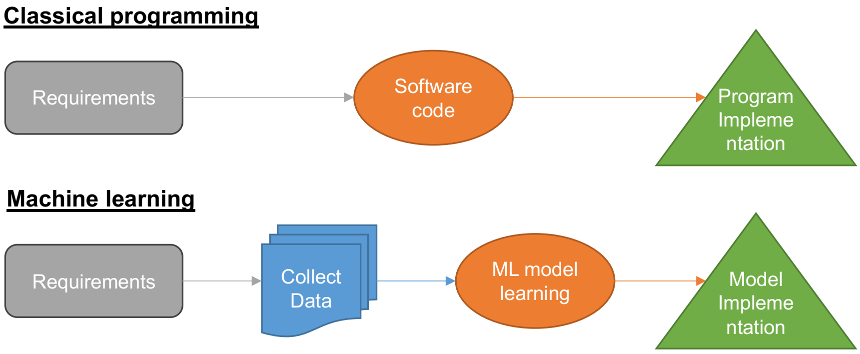

Figure 2-3: Main difference between ML and traditional programming (Source: own elaboration).

To learn the parameters of a ML model for a given application, the following choices need to be made:

I. Protocol: Select the approach and methodology for training the model (Data collection and learning strategy).

II. Dataset S: Determine the range of inputs and the dataset used for training and evaluation.

III. Hypothesis function space (h_θ): Choose the set of possible functions that the model can use to map inputs to outputs.

IV. Exploration algorithm of h_θ space: Select an algorithm to navigate the hypothesis function space and find an optimal or near-optimal solution.

V. Loss function and performance metrics: Decide on the appropriate loss function to measure the discrepancy between predicted and true outputs, as well as performance metrics to evaluate the model's effectiveness.

In the following section, it will be provided a classification of ML techniques based on how data is collected, the learning strategies employed, and the hypothesis used [14].

Classification elements of ML techniques

I. Data collection

The selection of a data collection protocol is a crucial factor in classification. To accurately describe a protocol, several criteria need to be considered, including the nature of data sources and sampling methods.

- **Passive vs active acquisition:** In passive acquisition, the data is collected independently of the ML decisions made during the learning phase. On the other hand, active acquisition refers to data collection that is dependent on the ML decisions taken during the learning phase.

- **Offline vs online learning:** Offline learning refers to the process of using a fixed or 'frozen' training set for learning. In contrast, online learning involves using data acquired during the operation or runtime of the system for learning purposes.

II. Learning strategies

One of the main criteria for AI classification is the learning strategy used. For ML, four learning methods are identified:

- **Supervised learning:** This strategy relies on labelled data to learn from. The algorithm compares the model's predictions with the known labels and adjusts its parameters to minimise the difference. The goal is to improve the model's accuracy by learning from the labelled data and making predictions on unseen examples. With a labelled training dataset $S = \{x_i, u_i\}$ with $x_i \in X$ (input space), and $u_i \in U$ (decision space), the system learns a function h_θ, with parameters θ, as close as possible to the target function, or distribution $h_\theta: X \to U$. U is discrete for classification problems and continuous for regression or density estimation problems (see Figure 2-4).

- **Semi-supervised**: This approach is applied when there is both labelled and unlabelled data available for training. The learning algorithm utilises the labelled data to measure the ML model's output against known labels, like supervised learning. Additionally, it leverages the unlabelled data to enhance the learning process and improve the model's performance. By making use of the unlabelled data, the algorithm aims to refine the model's accuracy and generalisation abilities beyond what can be achieved with only labelled data. The system learns using both a small amount of labelled data $\{x_i, u_i\}$ and a large amount of unlabelled data $\{x'_i\}$ in the training dataset S.

- **Unsupervised learning:** This method is employed when there is no labelled data set provided for training. The learning algorithm processes the available data set without any explicit knowledge of the correct labels. Instead, it focuses on uncovering patterns, structures, or relationships within the data. The algorithm utilises a loss function to assess the model's progress towards achieving a stable solution. Through iterative adjustments of the model's parameters, the algorithm aims to maximise the accuracy and utility of the ML model without relying on predefined labels. The system learns the underlying structure of data using an unlabelled dataset $S= \{x'_i\}$.

- **Reinforcement learning:** This strategy is used when an agent has access to an environment in which it can interact and learn through trial-and-error. The agent takes actions within the environment and receives rewards, either positive or negative, based on the consequences of its actions. These rewards serve as feedback to the learning algorithm, which updates the ML model's parameters accordingly. The goal is to optimise the model's performance by learning from the sequence of actions and rewards, enabling the agent to make informed decisions that yield favourable outcomes in the given environment.

The scope of this work is limited to passive, offline and supervised learning, as it seems to be a reasonable first step towards certification. Moreover, current efforts also define their scope in a similar way and aligning this work with them could provide valuable synergy.

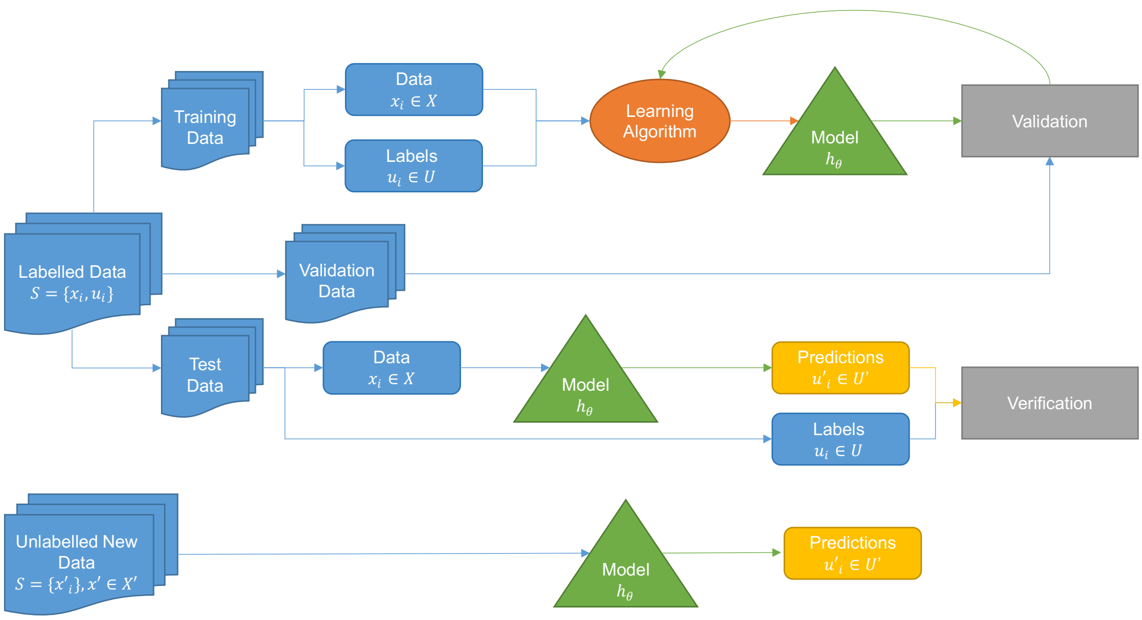

Figure 2-4: Supervised learning (Source: Own elaboration).

III. Hypothesis function space h_θ

Table 2-1 gives a short list of the most common ML techniques. Each technique relies on one or several hypothesis function space h_θ and one or several exploration algorithms. The underlined techniques correspond to those most used for supervised learning.

Techniques	Sub- Techniques	Applications
Linear models	Linear and logistic regressions, support vector machines (SVR)	Classification, regression
Neighbourhood models	KNN, K Means, Kernel Density	Classification, regression, clustering, density estimation
Trees	Decision trees, regression trees	Classification, regression
Graphical models	Naive Bayesian, Bayesian network Conditional Random fields	Classification, density estimation
Ensemble of models	Random Forest, Adaboost, Gradient Boosting	Classification, regression, clustering, density estimation
Neural Networks (NN)	Deep Learning (CNN, RNN...)	Classification, regression

Table 2-1: Taxonomy of ML techniques [14].

2.1.2 Levels of autonomy

AI is widely recognised as a crucial facilitator of system autonomy. When a device operates autonomously, it should be capable of assuming various tasks typically assigned to a human operator. A well-accepted framework for categorising autonomy is presented in [15]. Additionally, SAE International™ has established levels of driving automation in SAE J3016™ [16], while EASA in its Roadmap 2.0 has updated a classification for AI/ML applications in the aviation sector (Table 2-2).

Level AI 1: assistance to human	Level AI 2: human/machine teaming	Level AI 3: more autonomous machine
• Level 1A: Human augmentation • Level 1B: Human cognitive assistance in decision and action selection	• Level 2A: Human and AI-based system cooperation • Level 2B: Human and AI-based system collaboration	• Level 3A: The AI-based system performs decisions and actions, overridable by the human. • Level 3B: The AI-based system performs non-overridable decisions and actions.

Table 2-2: EASA Classification of AI/ML Applications [10].

SAE J3016™ focuses on categorising autonomy, while the EASA classification emphasises the role of AI within the overall system, including human interaction. In [17], a mapping is proposed between these classifications, which is presented in Table 2-3 and updated with EASA Roadmap 2.0.

AI Autonomy Level [15]	**SAE J3016™ Automation level** [16]	**EASA AI/ML Classification** [10]
1	0	1A
2	1	1A
3	1	1B
4	2	2A
5	2	2A
6	2	2B
7	3	3A
8	4	3A
9	4	3A
10	5	3B

Table 2-3: Mapping of AI/ML Levels from [17] **updated with EASA roadmap 2.0** [10].

Another approach commonly discussed in research papers about the concept of autonomy or human-machine teaming in military operations is the degree of human involvement in the operational Observe, Orient, Decide and Act (OODA)-Cycle, as referenced in [18]:

a) *'Human in the loop':* the human retains complete control over the autonomous system. The system may provide recommendations or suggestions, but the final decision-making authority rests with the human. The human can review and approve or reject the system's suggestions before taking any action (EASA classification equivalence: 1A, 1B & 2B).

b) *'Human on the loop':* the autonomous system can take actions on its own, but the human can intervene or override the system's decisions if necessary. The system operates autonomously most of the time but periodically involves the human for monitoring, intervention, or making high-level decisions (EASA classification equivalence: 2B & 3A).

c) *'Human out of the loop':* the system operates independently without requiring any input or intervention from the human. The system is designed to make decisions and take actions without human involvement. The human is not asked to confirm or approve any actions, nor can they easily abort or intervene in the system's operations (EASA classification equivalence: 3B).

2.1.3 Criticality

The criticality of a system in aviation, as defined in ARP4754 [19], is a measure of its importance and potential impact on safety and operational performance within the context of the entire aircraft or aviation system. Safety critical systems undergo more stringent development and certification processes to ensure their reliability and safety. Some systems may be safety critical in both civil and military contexts, while others may be safety critical only in the military domain. Autonomy, on the other hand, refers to the ability of a system or technology to perform tasks, operations, or functions without continuous human intervention. Autonomy can range from simple decision-making tasks to complex operations involving sensing, processing, analysis, and execution, all without direct human control. It's important to note that while autonomy can be a defining characteristic of a system, the criticality of a system is not solely dependent on its level of autonomy. For example, an avionics ISR system may have full autonomy, but if it is performing routine surveillance like gathering and classifying information, its safety criticality may be lower since it does not compromise the airworthiness of the air vehicle. On the other hand, a system may be safety critical, even if it involves a human-in-the-loop approach rather than full autonomy.

Imagine a military reconnaissance remote pilot aircraft (RPA), a state-of-the-art aerial vehicle designed for surveillance and intelligence gathering missions. This RPA is equipped with advanced technology and carries on board two distinct avionics systems:

a) ISR (Intelligence, Surveillance and Reconnaissance) system:

<u>Autonomy level:</u> Fully autonomous

<u>Safety critical:</u> Low

The ISR system is responsible for the routine surveillance and intelligence gathering tasks of the RPA. It operates autonomously, using a combination of sensors, cameras, and data processing algorithms to scan the terrain, identify objects of interest and transmit valuable data to the control centre. However, its functions are primarily data-centric and do not interfere with the airworthiness of the aircraft. In other words, the ISR system does not directly affect the RPA's flight control, making it less safety critical. Its primary mission is to provide real-time intelligence to military personnel.

b) Flight control system:

<u>Autonomy level:</u> semi-autonomous or non-autonomous

<u>Safety critical:</u> High

The flight control system, on the other hand, manages the actual operation and flight dynamics of the RPA. It involves human intervention in the form of a highly skilled pilot stationed on the ground who remotely controls the flight path and manoeuvres of the aircraft. Although the RPA can execute many flight tasks independently, such as maintaining altitude and following a predetermined flight plan, it relies on the pilot's expertise for tasks that require human judgement, such as adapting to changing weather conditions or avoiding unexpected obstacles. The pilot plays a crucial role in ensuring the safety of the RPA, which makes this system more safety critical despite its semi-autonomous nature process.

2.1.4 The W-shaped process

The W-shaped process, as illustrated in Figure 2-5, is a recently proposed software development life cycle (SDLC) that customises the conventional V-shaped software development process to better suit the unique requirements of ML-based (sub-) system development. The main objective of the W-shaped process is to effectively manage the machine learning development process, to ensure the performance of system outputs and behaviours, like what is normally achieved in traditional software development. It represents a paradigm shift from design assurance to learning assurance.

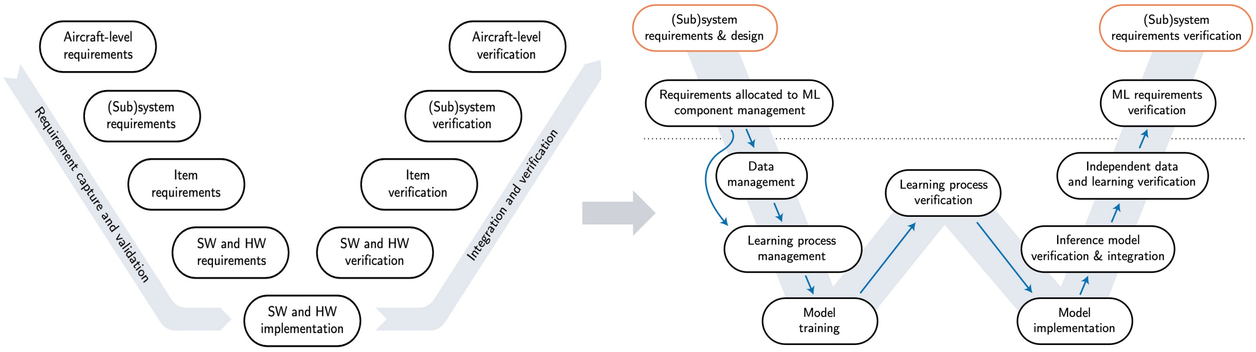

Figure 2-5: V-shaped design assurance vs W-shaped learning assurance processes [4], [5], [20], [21].

Getting into the different steps of the W-shaped learning assurance process:

- Requirements management: Requirements management and requirements verification (top right) are covered by traditional system development (indicated by the dashed line of the W-shaped development process).
- Data management: Datasets for training, validation, and evaluation are generated in alignment with the specified requirements. This process may encompass tasks such as data collection, annotation or labelling, and data processing.
- Learning process management: This stage comprises all the necessary components that precede the training phase, covering aspects such as defining metrics, establishing a strategy for model selection, evaluating various models or architectures, and configuring the software and hardware environment in which the actual training takes place.
- Model training and validation: During this iterative training and validation cycle, the objective is to identify the most optimal model, considering the architecture and hyperparameters that yield the best performance.
- Learning process verification: the single trained model undergoes evaluation using a test dataset. This evaluation encompasses assessing its generalisability, performance guarantees, and identifying instances of failure, all of which are subsequently considered in a safety assessment.
- Model implementation: This step encompasses all the necessary procedures for executing the model on the software and hardware platform, typically distinct from the development environment.
- Inference model verification and integration: Ensuring that the model meets performance guarantees in terms of accuracy based on the specified metrics and requirements. It also includes standard software verification elements such as execution time, memory usage, and other relevant factors.
- Independent data and learning verification: Step designed to bring closure to the data management life cycle. It assures that data has been correctly utilised throughout the process and aligns with the defined requirements, considering aspects such as completeness and representativeness.

2.1.5 AI applications in military avionics

In civil aviation, aircraft types are classified according to their category, maximum take-off mass and passenger capacity, although they all share a common purpose: transportation from point A to point B. Examples are normal, utility, aerobatic and passenger aircraft (CS.23) or large aircraft (CS.25). Today, military aviation encompasses various roles, including fixed-wing and rotary-wing aircraft, each specialised for specific tasks. Some aircraft are designed for multiple roles or can adapt their roles during missions, known as swing-role aircraft. These roles have evolved through years of combat experience and continue to serve in post-war stabilisation and peacekeeping efforts.

The adaptability of weapons, delivery methods, sensors, and avionics technology plays a crucial role in the versatility of military aircraft. While traditional roles persist, changing conflict dynamics may lead to the emergence of new roles or alternative approaches. These developments are largely driven by advancements in sensor and avionics technology, with improved sensor systems, stealth techniques, and onboard computing systems enhancing aircraft capabilities. Some essential roles needed in the realm of military defence are outlined below [22], [23].

I. **Combat aircraft:**

- **Fighter:** Fighter aircrafts are designed for air-to-air combat, emphasising speed and manoeuvrability to outmanoeuvre enemy aircraft. They perform both offensive and defensive roles, and their main weapons are air-to-air missiles and cannons. Some fighters also have ground attack capability, which has given rise to the concept of the multirole combat aircraft. These aircraft are intended to establish and maintain air superiority by denying enemy air power. They must respond quickly to threats, engage enemies beyond visual range and engage in dogfights if necessary. These aircraft require agility and advanced systems for navigation, target identification, weapons aiming and tactical communication. Modern fighters include the F-16, Rafale, Eurofighter Typhoon, Panavia Tornado, Su-30MKI, MiG-29 and Gripen.
- **Bomber:** Bombers, in general, are combat aircraft designed to attack ground and naval targets, employing a range of weapons from bombs to missiles. Strategic bombers carry powerful conventional or nuclear weapons for long-range missions against critical targets like supply bases. Examples include the B-2 Spirit and B-52.
- **Multirole combat aircraft (MRCA):** MRCAs are versatile aircraft capable of performing a variety of tasks such as air-to-air combat, bombing and aerial reconnaissance. The main reason for developing them is cost-effectiveness through a shared airframe. While specialised aircraft excel in specific roles, MRCAs can effectively perform multiple roles. For example, air superiority fighters, optimised for agility, can also attack land. Interceptors, designed for speed and close combat, can pursue, and neutralise enemy aircraft. Aircraft roles can be changed by modifying onboard armament or equipment, and some, such as the F-14, were upgraded for additional roles. The Eurofighter Typhoon, Panavia Tornado, and Dassault Rafale are examples of MRCA.

- **Electronic warfare (EW) aircraft:** An EW aircraft is a military plane equipped with EW systems designed to disrupt enemy radar and radio systems through radar jamming and deceptive tactics. EW technology involves detecting, identifying, and locating electromagnetic energy sources and employing self-protective jammers to block the opponent's access to the electromagnetic spectrum. EW operations can be conducted from land, sea, or air, employing various self-protective jamming methods like noise and frequency jamming, as well as countermeasure dispensing systems (CMDS). These capabilities enable the aircraft to operate stealthily by deceiving enemy radar systems. The F-35, for example, boasts advanced EW capabilities that allow it to reach heavily defended targets and neutralise threatening enemy radars.

- **Maritime Patrol (MPA):** An MPA is specifically designed for long-duration operations over coastal and water territories, with the primary mission of detecting, identifying and neutralising enemy ships and submarines using air-to-surface weaponry, torpedoes, and underwater mines. Equipped with advanced sensors such as sonar and radar, MPAs also play a crucial role in maritime search and rescue missions. Over 90 years, MPAs have evolved into highly complex aircraft capable of carrying out a variety of tactical, strategic, civil, and humanitarian tasks. They must move quickly to distant patrol areas, remain for long periods and have the endurance to return to base with fuel reserves. Typical tasks of MPAs include anti-surface unit warfare, anti-submarine warfare, search and rescue, exclusive economic zone protection, customs and excise cooperation, and various other roles in naval and police duties. Some examples include Lockheed P-3C, Lockheed S-3 Viking, Dassault Atlantic and Tupolev Tu-20.

II. Non-combat:

- **Military transport aircraft:** Fixed-wing and rotary-wing aircraft play crucial roles in military transportation operations, facilitating the movement of military personnel, cargo, and equipment for both routine and urgent military missions. These operations encompass a wide range of activities, such as rescue missions, tactical and strategic airlifts to locations with unprepared or semi-prepared runways. These transfers occur not only to forward bases but also to regions where commercial air services are not accessible. An example is the A400M.

- **Battlefield surveillance aircraft:** In military operations, having detailed real-time intelligence about the battlefield is crucial for commanders and planners. To meet this need, various commercial aircraft have been converted for this purpose, in addition to specialised military aircraft. These aircraft are equipped with radar systems positioned on the upper or lower part of the aircraft's frame, enabling them to scan the ground at an angle. They follow a predetermined flight pattern at a safe distance from enemy defences, identifying both stationary and moving targets. These findings are cross verified with information from other sensors or remote intelligence databases to create a comprehensive picture of the battlefield, including the positions and movements of both enemy and friendly forces. The mission crew collaborates as a team to construct this battlefield overview and can function as an airborne command centre to direct operations, such as air or ground strikes. Grumman E-2 Hawkeye; Boeing E-3 Sentry; Lockheed P-3 AEW; Tupolev Tu-126 AEW; Westland Sea King

- **Photographic reconnaissance aircraft:** Photographic imagery intelligence (IMINT) plays a vital role in confirming SIGINT data by providing high-resolution imagery that verifies and enriches the information collected. These images can reveal troop numbers, formations, battalion markings, and force deployments. Obtained via satellites or aircraft flying at different altitudes, IMINT activities carry risks such as loss of aircraft due to missile attacks or diplomatic tensions, as seen in historic incidents such as the U-2 aircraft over the Soviet Union in 1960 and the case over Cuba in 1962. IMINT-equipped aircraft employ advanced technology for optimal image quality, including mission computers, altitude speed optimisation, lens aperture adjustments and high-speed focal plane shutters. The resulting tactical information, combined with other camera systems, provides a real-time view of the battlefield and confirmation of damage. Canberra PRY; e Jaguar GR1; e Tornado GR4.

- **Airborne early warning and control aircraft (AEW&C):** Timely detection and warning of an impending airborne attack is crucial for ensuring air superiority and providing defensive forces with ample time to prepare a robust defence. Furthermore, it is essential to alert ground and naval forces about the approaching attack, allowing them to take appropriate defensive, evasive, or countermeasures actions. The airborne platform is outfitted with advanced radar and electronic intelligence capabilities through sensors to effectively monitor communication and radar emissions. Additionally, it incorporates an electronic support measure (ESM) system designed for the long-range detection of aircraft, ships, and vehicles, while also serving as a command-and-control centre for battle management (C2BM) during aerial engagements by directing fighter and attack aircraft strikes. When flying at higher altitudes, the aircraft's radar and other sensors have improved visibility, enabling the on-board processor to detect, identify, and track both ground and slow-moving targets. Consequently, AEW&C units are responsible for surveillance and act as C2BM by utilising command data links. To ensure their safety, AEW&C aircraft are equipped with EW systems and self-protective jammers to minimise vulnerability to counterattacks. While ‘Airborne Warning and Control System’ (AWACS) is the specific term used for the system installed in E-3 and Japanese Boeing E-767 aircraft, it is often used more broadly as a synonym for AEW&C.

- **Air-to-air refueler:** Aerial refuelling, a widely adopted practice in modern aviation, particularly in military operations, involves the transfer of aviation fuel from one aircraft (the tanker) to another (the receiver). This method enables the receiver aircraft to take off with a greater payload, such as cargo, or personnel, while carrying less fuel to adhere to aircraft maximum weight limitations. After becoming airborne, the receiver aircraft can then receive additional fuel from the tanker, thereby extending its operational range or endurance, making it capable of loitering over a target location for an extended period. Both the receiver aircraft and the tanker are equipped with a precision probe and drogue system for this purpose, with the receiving aircraft featuring a fuel probe in the front section, and the tanker being fitted with a drogue to facilitate fuel transfer. An example is the MRTT A330.

In terms of architectural platform, compared to civil aircraft, military aircraft have specialised mission and combat aircraft systems, as well as weapons systems. In the following Figure 2-6, it is presented two examples of architecture: one of a combat aircraft and the other of a non-combat aircraft (from left to right). Both present the typical avionics, hydraulics, propulsion systems and specific to military aircraft, the mission related systems. In the air superiority example, for example a fighter, there is no mission crew, only the flight crew, which displays the mission information on their heads-up display (HUD) or head-mounted display (HMD). In the case of battle surveillance, there is a mission crew that analyses all the information received through specific mission displays. Another main difference is that fighter aircraft have a weapon system and non-combat aircraft do not. Both are equipped with defensive aids systems (DASS).

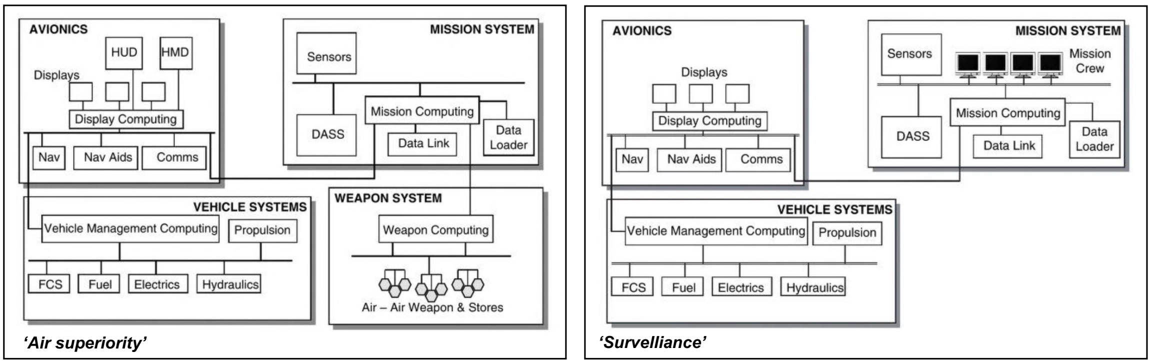

Figure 2-6: Typical platform architecture comparison; air superiority vs surveillance [22].

Avionics refers to a complex structure made up of various electronic subsystems, each exhibiting typical system characteristics. This system consists of multiple major subsystems that work together to fulfil its overall function. These major subsystems can further be broken down into minor subsystems or equipment, which must operate and interact harmoniously to support the functioning of the entire system. Each of these minor subsystems relies on the proper functioning of its components or modules to contribute to the overall effectiveness of the system. This hierarchical arrangement resembles a pyramid, where the integrity of the entire system relies on the successful operation of its lower-level components. Figure 2-7 presents a breakdown structure of the typical avionics found in military aircrafts. Some are common with civil aircraft, such as Traffic Collision Avoidance system (TCAS) or autopilot, and others are military specific, such as DASS or HUD.

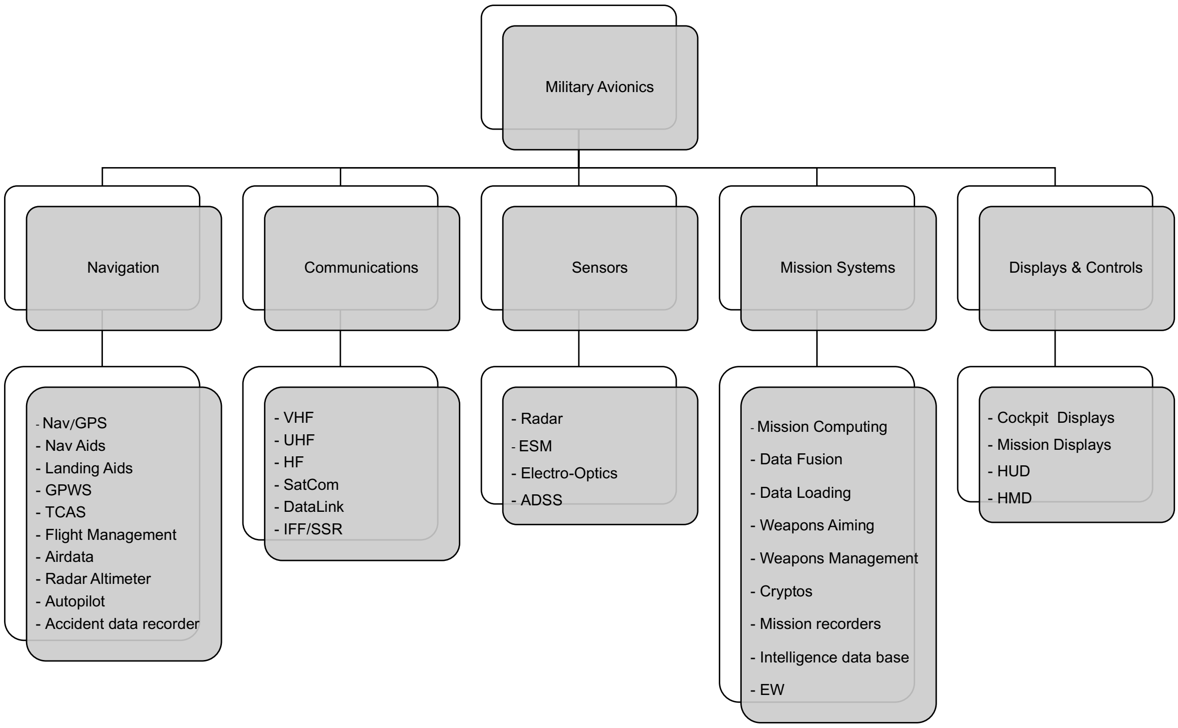

Figure 2-7: Military Avionics [22].

While businesses and academia have been at the forefront of AI development, the defence sector in general has been slower to adopt these technologies. However, as AI technology continues to evolve, it is expected to play an increasingly critical role in reshaping the future of avionics across military domains by introducing autonomous systems, enhancing threat detection, optimising mission planning, providing decision support, improving situational awareness, training, and air traffic management. AI has demonstrated its ability to outperform humans not only in routine military tasks, but also in more complex tasks. For example, AI has excelled in tasks such as aerial combat, demonstrating its potential to improve the accuracy of data analysis, and provide early warnings of threats [6]. AI possesses remarkable processing speed, a critical attribute as it becomes integrated into military decision-making processes. In these situations, where the military may have mere minutes to respond to a threat, AI's rapid data processing capability is indispensable. Moreover, AI algorithms can be integrated into automated systems and elevate them, especially those that perform tasks considered monotonous, dangerous, or undesirable to humans. Leveraging modern AI capabilities, allows for greater autonomy, moving beyond traditional automation to fully autonomous systems. This transition allows soldiers to avoid risking their lives in activities such as mine clearance or dangerous intelligence, surveillance, and reconnaissance operations. Autonomous vehicles can tirelessly patrol areas for long periods without rest, process data much faster than human analysts, thus allowing military personnel to focus on other critical responsibilities [6].

In terms of military avionics, the incorporation of AI into these fields will greatly enhance the effectiveness and capabilities of alliance military air forces. The next few subsections will briefly outline a few of the most important areas of technological progress in AI that have potential application [1], [6]–[8], [24], [25].

- **Autonomous Flight:** Flight guidance, navigation, and control employ AI techniques, specifically machine learning, to implement intricate guidance and flight control behaviours for swarms of platforms. An example is the development of unmanned aerial vehicles (UAVs) or drones capable of autonomous flight and autonomous navigation enabling extended flight durations and accessing dangerous or hard-to-reach areas. Presently, most UAVs are not entirely autonomous but rather remotely piloted (RPAS), primarily due to the limitations of current technology (conventional automated software) in ensuring secure flight formations.

- **Intelligence, Surveillance, and Reconnaissance (ISR):** AI/ML systems can analyse vast amounts of sensor data, including radar and infrared signatures, to detect and classify potential threats such as enemy aircraft, missiles, or drones. These systems can perform ISR acquisition without human intervention. This capability enables early warning systems, target prioritisation, pattern recognition, electronic warfare (EW), and effective deployment of countermeasures.

- **Mission Planning and Execution (MPE):** encompasses two phases within a military mission: initial pre-planning and real-time adjustments during the execution phase. AI/ML algorithms assist in generating optimal flight routes, allocating resources, and optimising fuel consumption. These systems consider factors such as weather conditions, terrain, enemy positions, and available resources to enhance operational effectiveness and efficiency. These plans consider various factors and provide recommendations to human operators, leading to more informed and effective decision-making.

- **Autonomous Weapon Capability:** Target detection and identification encompasses the use of AI technology to detect and identify potential targets within the Find and Fix activities of the F2T2EA-Cycle (Find, Fix, Track, Target, Engage and Asses). Using advanced sensors, AI algorithms analyse real-time data to select and execute missions with minimal human intervention. In this F2T2EA cycle, the most ethically controversial activities to automate are targeting and engaging. AI algorithms must evaluate target characteristics, assess collateral damage risks, and predict possible outcomes to ensure accurate and ethical decision making. Autonomous weapon systems must adhere to predefined rules of engagement. Threat Assessment and Aiming Analysis (TAA) can implement AI fir threat analysis and weapon aiming.

- **Decision Support:** AI/ML systems can provide real-time analysis and decision support to military aviators. They process sensor data, intelligence reports, and historical data to generate insights, assess potential threats, and recommend appropriate actions, leading to faster and more informed responses in dynamic operational environments. Another example could be the air-to-air refuelling system. The operation of inserting the refuelling hose into the receiver intake during in-flight refuelling is another critical area where computer vision technology holds immense potential. This intricate procedure requires precision and coordination, as the tanker operator must precisely guide the hose to the receiver intake port while both aircraft are in motion. Computer vision systems, equipped with advanced cameras and image recognition AI algorithms, have the capability to provide guidance and even automate this process. By providing real-time visual information, computer vision can help tanker operators accurately align the refuelling hose with the receiver intake, considering variables such as aircraft movement, turbulence, and changes in relative positions.

- **Command, Control, Computers, Communications, and Intelligence (C4I):** AI/ML techniques are utilised to perform data fusion and intelligence analysis. They fuse data from multiple sources, such as radar, satellites, and ground-based sensors, to create a comprehensive situational awareness picture. This integrated data enables accurate threat assessments, target identification, and intelligence analysis. This heightened situational awareness allows for better mission execution and increased operational effectiveness.

- **Cyber Security and Resilience (CSR):** CSR can use AI techniques, specifically data analytics, to detect anomalies and adversarial activities in cyberspace. AI/ML systems can rapidly detect and analyse potential threats, establish behaviour patterns to identify anomalies, monitor networks in real time, and automate responses to contain attacks. AI's predictive analytics foresee future threats, while vulnerability assessments and phishing detection enhance defences. It swiftly recognises new malware strains and adapts defences against evolving cyber adversaries. By integrating AI, military cybersecurity gains proactive readiness and resilience against diverse cyber threats. Nevertheless, it is meaningful to note that these resilience avionic systems are not addressed in this scientific work.

- **Training and Simulation:** AI/ML systems can be used to develop advanced flight simulators and virtual training environments. These systems can generate realistic scenarios, adapt to pilot or system operator performance, and provide personalised training experiences, improving pilot and system operators' skills.

- **Predictive Maintenance:** AI/ML algorithms can analyse aircraft sensor data to predict and prevent equipment failures before they occur. By monitoring parameters such as engine performance, temperature, and vibrations, these systems can identify anomalies and schedule proactive maintenance, minimising downtime and enhancing aircraft availability. In-flight monitoring systems are considered airborne systems as well. This section does not refer to the Reduced Life Cycle Cost (RLC) which can also use AI techniques to analyse big data in the perimeter of production, maintenance, logistics and overall system lifecycle.

- **Air Traffic Management (ATM):** AI/ML systems contribute to improving air traffic management by optimising airspace utilisation, reducing congestion, and improving flight safety. These systems can predict and mitigate potential conflicts, optimise routing, and facilitate more efficient coordination among military aircraft and civilian air traffic. Also not included in the scope of this thesis since it does not refer to airborne.

The Appendix A, Table 7-1: Classification of military AI-based airborne systems provides an own elaborated comprehensive classification of AI/ML-enabled systems for the previously described military applications (inspired by the 'Classification for Avionics Capabilities enabled by Artificial Intelligence,' [17]).

2.2 Regulatory framework: Contrasting Civil and Military Airworthiness

Alignment between military and civil airworthiness is an essential factor in standardising the implementation of AI/ML systems in military avionics. Before presenting the challenges of introducing AI/ML systems on military aircraft and whether they could be aligned with the approach being undertaken in the civil sector, this paper briefly outlines some key differences between civil and military regulations in the field of airworthiness.

2.2.1 Regulatory framework

Military and state aircraft (such as custom, police, search and rescue, firefighting, coast guard, or similar activities or services) are excluded from fundamental civil aviation agreements and laws, such as the Chicago Convention of the International Civil Aviation Organisation (ICAO) (Article 3, [26]) or the EASA Basic Regulation (Article 2, [27]), worldwide and European aviation regulations respectively.

> *“This Regulation shall not apply to:*
>
> *(a) aircraft, and their engines, propellers, parts, non-installed equipment and equipment to control aircraft remotely, while carrying out military, customs, police, search and rescue, firefighting, border control, coastguard or similar activities or services under the control and responsibility of a Member State, undertaken in the public interest by or on behalf of a body vested with the powers of a public authority, and the personnel and organisations involved in the activities and services performed by those aircraft;*
>
> *(b) aerodromes or parts thereof, as well as equipment, personnel, and organisations, that are controlled and operated by the military;” [Article 2, [21]]*

Civil or commercial aircraft must undergo certification by Civil Airworthiness Authorities (CAAs) to adhere to international aviation regulations. On the other hand, military aircraft, being state-owned, undergo their “self-certification" process tailored to specific airworthiness criteria, in which the government ensures that the aircraft is airworthy (‘certification’) and meets the performance requirements (‘qualification’) of its controlled military airspace. If these military aircrafts operate within their designated segregated airspace, there is no conflict with civilian aircrafts [28]. The specifics regarding equipment qualification, operational standards, and ATM matters for military aircraft are classified and not bound by international airspace utilisation rules. Moreover, military aircraft not only have restrictions on flying in civil airspace but also require special permission to fly in foreign segregated airspace from the state that has sovereignty over that airspace. In other words, military aircraft have control over their operations only within their own national airspace, which is segregated at the same time from e civil airspace, regulated by international laws. The operational risk associated with these aircraft is defined and accepted by the respective service, although the level of risk may vary depending on the purpose and type of the aircraft. Military aviation always must strike a balance between safety and mission accomplishment. This is related to the fact that, since military aircraft are intended to overfly enemy territory in hostile airspace, they carry weapons or operate from airfields without infrastructure. In short, military aircraft must be as reliable as possible in the worst environmental conditions.

Defining airworthiness criteria and certification basis for a new program is a crucial task that is closely tied to pre-contractual activities. The absence of a general regulatory framework leads to program-specific procedural systems, and discussions with relevant authorities can extend over a significant time frame, considering both domestic and international considerations. While the industry attempts to incorporate common elements across projects, the outcome usually results in a program-specific procedural system. The duration of pre-contractual discussions to determine the appropriate certification and airworthiness path can span from several months to years, depending on the complexity of the product. For export projects, additional discussions take place between the industry and various authorities involved in overseeing the certification and airworthiness steps. This includes engaging with customer authorities, the authorities of the design and manufacturing country, and potentially the EASA if their contribution is relevant to the product. Other government agencies may also be involved when qualification is impacted. Bringing together all stakeholders involved in the airworthiness process is primarily the industry's responsibility, and achieving this task in a short time frame is often challenging.

It is quite clear that airworthiness certification of military aircraft cannot work in the same way as civil airworthiness [29]. However, as dedicated military airspace becomes scarcer, there is a growing need to establish a common set of standards to ensure compatibility and cooperation between military and civil aircraft within the national and international airspace. Lately, there is an increasing recognition by the military world, at least in Europe, that following the basic principles behind the civil regulations can be a benefit also in military aviation.

2.2.2 Airworthiness stakeholders and regulatory material

Examples of airworthiness stakeholders may include regulatory authorities, aircraft manufacturers, operators, standardisation organisations and maintenance repair and overhaul (MRO) organisations. For the initial implementation and certification of this new AI-based avionics systems, a deeper insight into the regulatory authorities, and aviation industry standardisation associations is made.

<u>In civil aviation</u>

ICAO establishes worldwide standards and recommended practices (SARPs), including safety standards for civil aviation. ICAO member countries, including EU countries, adopt these standards and incorporate them into their national regulations. As the executive body of the EU, the European Commission takes the ICAO standards as a reference and based on the needs and priorities of the EU, may propose additional regulations and directives to enhance safety and efficiency in aviation within the European context. On the other hand, EASA is responsible for implementing and enforcing EU regulations concerning aviation safety. The agency works closely with the European Commission and EU member states to ensure harmonisation and effective implementation of the standards throughout the region. The European Commission regulation (EC) No. 748/2012 [30] encompasses a global framework that includes Part 21 regulations in Europe, which establish requirements for aircraft certification, covering design, production, and approval of modifications. Part 21 of the aviation regulations comprises Certification Specifications (CS), which provide detailed technical guidelines for aircraft design and certification. These processes are accompanied by Acceptable Means of Compliance (AMC) documents, which provide practical guidance for meeting the requirements effectively, and guidance material, such as industry standards that provide additional insights and best practices.

Specifically, CS 25.1309 and AMC 25.1309 pertains to “Equipment, systems and installations” for large aeroplanes. Both the International Society of Automotive Engineers (SAE) and the European Organisation for Civil Aviation Equipment (EUROCAE), representing the United States and Europe, respectively, play essential roles in advancing the aviation industry by providing technical expertise, promoting collaboration, and establishing common industrial standards that contribute to the safe, reliable, and efficient operation of aircraft and aviation equipment. RTCA DO-178C “Software Considerations in Airborne Systems and Equipment Certification” [31] is the latest iteration of the FAA avionics software standard, which is mandatory for all commercial airborne software. This standard plays a crucial role in ensuring flight safety by establishing a high level of confidence that the software performs its intended functions as defined in the system requirements. However, in the early 2000s, aviation authorities recognised that avionics safety was influenced by both software and hardware components. The importance of hardware in aviation safety became apparent and the DO-254 “Design Assurance Guidance for Airborne Electronic Hardware” was developed to establish consistent certification requirements for hardware. On a broader scale, SAE ARP4754A “Guidelines for Development of Civil Aircraft and Systems” [19] is applicable to civil aircraft and systems, while SAE ARP4761 “Guidelines and Methods for Conducting the Safety Assessment Process on Civil Airborne Systems and Equipment” [32] (including the new version ARP4761A) provides detailed safety analysis guidelines for these systems. This collective effort reflects the aviation industry's commitment to maintaining high levels of safety and harmonising certification processes for software and hardware components of airborne systems [33]. The equivalent standards of ARP4754A, ARP4761, DO-178C and DO-254 adopted by EUROCAE in Europe are ED-79A, ED-135, ED-12C and ED-80, respectively.

In the European Union, each member of the community has its own national civil airworthiness authority representing EASA. These national authorities work closely with EASA to ensure harmonisation and effective implementation of safety rules throughout the region. Examples of EU countries and their national civil airworthiness authorities include:

I. Germany: The national civil airworthiness authority in Germany is the Luftfahrt-Bundesamt (LBA).
II. France: The national civil airworthiness authority in France is the Direction Générale de l'Aviation Civile (DGAC).
III. United Kingdom: Prior to its exit from the European Union, the United Kingdom's national civil airworthiness authority was the Civil Aviation Authority (CAA). Following the UK's exit from the EU, the CAA continues to play a role in the regulation of civil aviation in the UK, although there are now regulatory differences between the UK and the EU.

In the United States, the equivalent of the European Union Aviation Safety Agency (EASA) is the Federal Aviation Administration (FAA). The FAA is the regulatory authority for civil aviation in the United States and is responsible for establishing regulations and safety standards for civil aviation in the country.

In military aviation

As mentioned before, ICAO sets SARPs to ensure uniform safety regulations for civil aviation worldwide. There is currently no equivalent international organisation that standardises military aviation regulations in a similar manner. While civil aviation in Europe is organised under the leadership of EASA as a coordination centre, the military aviation community in Europe does not have a single coordination centre. The European Defence Agency (EDA) has developed a regulatory framework for harmonising requirements among European militaries. In November 2008, the EDA Board assigned the European Defence Agency (EDA) the task of creating a forum known as the Military Airworthiness Authorities Forum (MAWA) [34]. In addition, they approved a roadmap for military airworthiness aimed at achieving harmonisation and standardised certification processes [35]. The goal of the MAWA Forum is to develop a comprehensive set of European Military Airworthiness Requirements (EMAR) [36]. It is important to note that compliance with EMARs is not mandatory, and each country retains the discretion to decide whether or not to adopt them. The growing importance of this issue is evidenced by the agreement signed in June 2013, which established enhanced cooperation between the EDA and the European Union Aviation Safety Agency (EASA). This agreement focuses mainly on the harmonisation of military aviation safety requirements, in particular in the context of airworthiness [37].

However, these efforts are voluntary and mainly limited to adoption within their respective regions, particularly within the European Union. Each military aviation regulator has established its own unique system of airworthiness regulations. The lack of a recognised method for unbiased assessment and comparison of these diverse regulatory frameworks presents a significant challenge when seeking acceptance between different military regulatory authorities. Every nation counts with a National Military Aviation Authority (NMAA).

Some examples are:

I. United States: In the United States, military aviation is overseen by the U.S. Department of Defence (DoD), which includes various branches such as the U.S. Air Force, U.S. Army Aviation, U.S. Navy, and U.S. Marine Corps Aviation. Each branch has its own authority responsible for regulating and overseeing its aviation operations.

II. United Kingdom: In the United Kingdom, military aviation is regulated by the Military Aviation Authority (MAA), which is an independent organisation responsible for ensuring the safe and effective operation of military aviation.

III. Germany: In Germany, military aviation is overseen by the Luftfahrtamt der Bundeswehr (LufABw), which translates to the "Aviation Office of the Federal Armed Forces." It is responsible for airworthiness, safety, and regulatory oversight of military aviation operations.

Figure 2-8 shows the interrelationship between German military aviation stakeholders as an example of organisations involved in national airworthiness. The relationship between these entities has been extracted from Deutschland Military Airworthiness Requirements (DEMAR) [38] can be summarised as follows:

Airbus Defence and Space (Airbus D&S), as an aerospace and defence company, may be one of the suppliers of military aircraft and related systems to the German Armed Forces, including the Luftwaffe. The Luftwaffe operates and maintains military aircraft and defines the operational requirements and capabilities needed for its missions. LufABw oversees the airworthiness, safety and regulatory compliance of military aircraft manufactured by Airbus D&S. It works closely with the Luftwaffe to ensure the smooth operation and safety of aircraft. The BAAINBw plays a role in the procurement and management of military aviation equipment and holds the Military Type Certificate (MTC), unlike civil aviation, where the type certificate is held by the manufacturer.

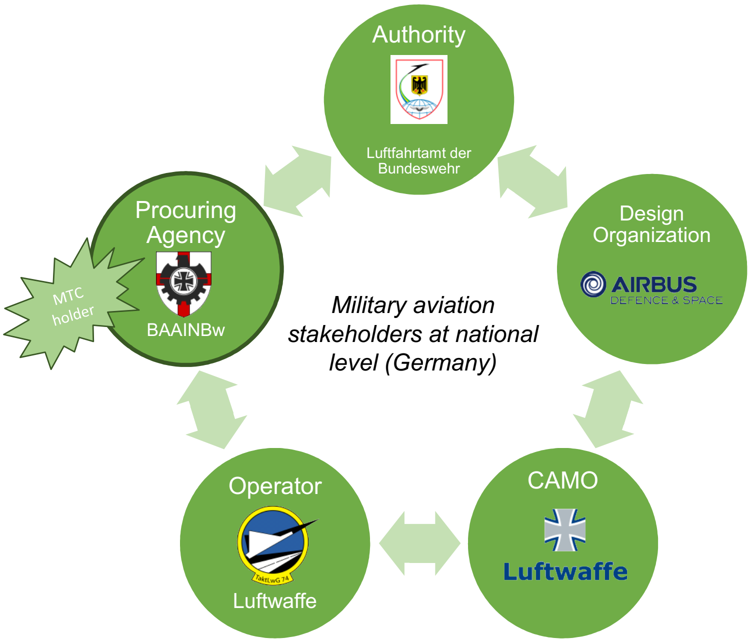

Figure 2-8: German military aviation stakeholders [Source: Own elaboration].

In military aviation, several international defence organisations play significant roles in coordinating operations and developing military aviation standards. NATO, a political and military alliance of member nations from North America and Europe, develops interoperability standards, called Standardisation Agreements (STANAGs), which are documents that establish common technical and procedural standards among member nations, facilitating interoperability in areas such as equipment, communication, and operational practices [39]. In general, STANAGs are developed for a wide variety of land, sea, and air defence applications.

Some examples of STANAGs related to avionics are:

- STANAG 4626 “Modular and open avionics architectures*”* [40], which outlines a series of open architecture standards designed for avionics architecture, with a particular focus on integrated modular avionics (IMA).
- STANAG 4452 “Guidance on software safety design and assessment of munition-related computing systems” [41].

On the other side, the DoD also established the defence standards (DOD-STDs & MIL-STDs) established by to implement uniform requirements and ensure interoperability for the military and its secondary industries, focusing explicitly on technical and engineering requirements, processes, procedures, practices, and methods. For many years, military organisations have developed both hardware and software using specialised standards such as DOD-STD 2167A “Defence Systems Software Development" [42] , MIL-STD 498 “Military Standard Software

Development and Documentation" [43] and MIL-STD 882e "System safety" [44]. This divergence from commercial standards was motivated by the belief that military applications had unique requirements, primarily focused on mission success. Currently, however, there is a rapidly growing trend towards convergence of military and commercial avionics standards. This convergence involves the widespread adoption of standards such as DO-178C and DO-254 on a global scale, along with the incorporation of new safety guidelines such as ARP4754A and the updated ARP4761A. Today, a wide range of aircraft, including fighters such as the Joint Strike Fighter and T-50, cargo, and refuelling aircraft such as the C-130, C-17, A400M, KC-46, as well as the UAVs or RPAS, are requiring compliance with DO-178C. In addition, increasing emphasis is being placed on compliance with DO-254 and the guidelines outlined in ARP4754A/ARP4761. This change represents a significant move towards harmonisation of military and commercial aviation standards and practices [33]. DO-178C is the latest iteration of the FAA avionics software standard, which is mandatory for all commercial airborne software. This standard plays a crucial role in ensuring flight safety by establishing a high level of confidence that the software performs its intended functions as defined in the system requirements.

The main differences between civil and military regulatory frameworks are highlighted in

Table 2-4:

Aspects	**Civil Aircraft Certification**	**Military Aircraft Certification**
Business Outlook	International Business through common rules and collaboration	Military aviation is under each individual nation sovereignty
Design Standard	Internationally accepted airworthiness standards	Different standards followed by different nations
Compliance	Compliance to Airworthiness Certification Criteria	Demonstration restricted to budget provisions
Certification Objectives	Verification and demonstration of compliance to regulatory requirements	Qualification and issue operational clearance for introduction into service
European Aviation Authority	EASA	None (*MAWA as future authority)
National Aviation Authority (Germany)	NCAA (LBA)	NMAA (LufABw)
Regulatory Material (software specific)	Part 21 + CS + AMC + GM + CRIs	National Regulations + EMAR 21 + MCRIs
Industrial Standards (software specific)	ARP4754A + ARP4761 + DO-178C + DO-254	MIL-STD 498 + MIL-STD 882E + STANAGs

Table 2-4: Regulatory framework summary.

2.2.3 System safety approach

System safety requirements are outlined in SAE ARP4761 [32] for civil aircraft and in the US DoD document MIL-STD 882E [44] for military aircraft. These standards define different risk factors ranging from “catastrophic damage”, representing potential loss of life and property, to the lowest risk factor “minor”, indicating minor inconvenience. The occurrence frequency is categorised as “frequent or probable” (1 in 10) to “extremely improbable” (1 in 10^{-8} for military and 1 in 10^{-9} for civil aircraft design). To assess safety, both standards use a hazard index, calculated by multiplying the damage severity with the frequency of occurrence. Design safety acceptance criteria are based on this hazard index.

These distinctions arise from the acceptance of a somewhat higher level of risk in military aircraft design to prioritise mission accomplishment. Nonetheless, both approaches adhere to the fundamental principle of airworthiness: the probability of an event should inversely correlate with the potential severity of its impact [28] (see Table 2-5). Some platforms such as the A400M, issue a Military Certification Review Item (MCRI), the equivalent of a Certification Review Item (CRI) in the military domain, which refers to a specific item or criterion that is evaluated during a certification review process. For example, the MCRI that addresses the introduction of MIL-STD-882D into the certification basis is called MCRI F-06 “Equipment, Systems and Installation (Safety)” [45]. This is done because the EASA requirements must be adapted to consider the specificity of a military aircraft performing military operations.

Frequency of occurrence [P per Fh]	FC Classification			
	Catastrophic	Hazardous	Major	Minor
$P>10^{-3}$ ‘Probable’				
$10^{-4}<P\leq10^{-3}$ ‘Probable’				
$10^{-5}<P\leq10^{-4}$ ‘Remote’				
$10^{-6}<P\leq10^{-5}$ ‘Remote’				
$10^{-7}<P\leq10^{-6}$ ‘Extremely Remote’				
$10^{-8}<P\leq10^{-7}$ ‘Extremely Remote’				
$10^{-9}<P\leq10^{-8}$ ‘Extremely Improbable’				
$P\leq10^{-9}$ ‘Extremely Improbable’				

Table 2-5: Failure Condition Level for military operations [44].

	Unacceptable area
	Area of exceptional deviation with approval of the Authority
	Area for military operations only
	Acceptable area

Table 2-6: Colours assignment.

The military safety objective for a given standard military operation phase is 1x10^{-6} Per Flight Hour at aircraft level.

Long before an aircraft is built, all these failure conditions are identified, and appropriate design measures are taken, and the quality of the design is assured. This process helps to define the system architecture, the number of control and power channels, the level of redundancy, etc. It also specifies a level of design assurance based on what the effects of a failure might be. To address this, DO-178C is divided into five levels ranked from Level A (the highest level of criticality) to Level E (the lowest level of criticality). These levels are formally referred to as "Developmental Assurance Levels (DALs)". Each avionics system is assigned one or more criticality levels based on a comprehensive system safety assessment, which evaluates the potential impact of each system on the overall safety of the aircraft. Depending on the safety criticality of the application, and on the aviation domain, an assurance level is allocated to the AI-based (sub)system too. Consequently, all hardware and software components of a given system must meet or exceed the assigned criticality level. As the criticality level rises from A to E, there is a corresponding increase in the stringency of requirements for documentation, design, review processes, implementation, and verification procedures.

Table 2-7 shows this classification:

Software Level Definition	Failure Conditions
Level A (DAL A)	Software whose anomalous behaviour, as shown by the safety assessment process, would cause or contribute to a failure of system function resulting in a **CATASTROPHIC** failure condition for the aircraft
Level B (DAL B)	Software whose anomalous behaviour, as shown by the safety assessment process, would cause or contribute to a failure of system function resulting in a **HAZARDOUS** failure condition for the aircraft
Level C (DAL C)	Software whose anomalous behaviour, as shown by the safety assessment process, would cause or contribute to a failure of system function resulting in a **MAJOR** failure condition for the aircraft
Level D (DAL D)	Software whose anomalous behaviour, as shown by the safety assessment process, would cause or contribute to a failure of system function resulting in a **MINOR** failure condition for the aircraft
Level E (DAL E)	Software whose anomalous behaviour, as shown by the safety assessment process, would cause or contribute to a failure of system function with **NO EFFECT** on aircraft operational capability or workload. If software component is determined to be Level E and this is confirmed by the certification authority, no further guidance contained in the DO-178 applies.

Table 2-7: Software level definition by DO-178 [31].

2.2.4 Harmonisation civil and military certification

For export projects, additional discussions take place between the industry and various authorities involved in overseeing the certification steps. This includes engaging with customer authorities, the authorities of the design and manufacturing country, and potentially the EASA if their contribution is relevant to the product. Other government agencies may also be involved when qualification is impacted. Bringing together all stakeholders involved in the airworthiness process is primarily the industry's responsibility, and achieving this task in a short time frame is often challenging. Lately, there is an increasing recognition by the military world, at least in Europe, that following the basic principles behind the civil regulations can be a benefit also in military aviation.

In the context of setting the initial certification basis of military aircraft, there are traditionally two main stages: Certification and Qualification. The first consists of demonstrating that the aircraft is safe for flight. It ensures that the aircraft is airworthy and meets specified safety standards and requirements. Qualification is the step that demonstrates that the aircraft performs as specified in a contract and ensures that the aircraft can meet the agreed capabilities required for military operations.

In the past, with legacy platforms, such as Eurofighter Typhoon, certification and qualification requirements were often mixed. Today, there is a growing trend towards a three-stage process. First, the civil certification for operations in civil airspace, usually for cargo transport or other non-military purposes. It ensures that the aircraft meets the required safety standards for operations in civil airspace. Secondly, the military certification for military operations, which often involves taking existing civil certification (if any) and adding additional requirements specific to military operations. This is done to ensure the safety of the aircraft in military contexts. Finally, the qualification: This stage to confirm that the aircraft can meet the specific capabilities required for military operations. It may involve additional tests or modifications beyond those contemplated in the military certification.

It's becoming increasingly questioned whether it's justifiable to invest in designing and certifying the same aircraft multiple times to meet the requirements of two distinct certification authorities in the case that the aircraft must fly in both airspaces. For future aircraft programs, the objective is to conform to the EMARs developed by MAWA. This would streamline the process and avoid double certification. Instead, EASA would recognise authorities and regulatory frameworks. Essentially, this means that if an aircraft meets EMARs regulations, it should be accepted by both civil and military authorities without the need for redundant certification processes. This approach is intended to increase efficiency and reduce the burden of certifying new aircraft.

A400M ‘dual certification’

For instance, the A400M is a highly versatile military transport aircraft designed to perform a variety of roles and functions. In its logistics transport role, it offers impressive capabilities, such as long-range travel with in-flight refuelling, high cruise speeds, a spacious cargo hold with substantial payload capacity, a flexible cargo handling system and the ability to perform medical evacuations. In its tactical transport role, the A400M excels in low-speed manoeuvres, short take-offs, and soft field operations, making it suitable for demanding tactical missions. It also has autonomous ground operation capabilities and can perform aerial deliveries of paratroopers and cargo. The aircraft prioritises survivability, which increases its effectiveness in demanding operational scenarios. The A400M is not limited to transport roles; it also serves as an in-flight refuelling platform. It is equipped with two or three refuelling systems and operates effectively at a wide range of altitudes and speeds [46].

The A400M program began with the A400M Development and Production Phase Contract, awarded to Airbus in 2003 by OCCAR (Organisation for Joint Armament Cooperation). This contract represents several participating states, including Belgium, France, Germany, Luxembourg, Spain, Turkey, and the United Kingdom, reflecting the international collaboration involved in the program. OCCAR plays a crucial role in the management of cooperative defence equipment projects [46]. To ensure the airworthiness and qualification of the A400M, a collaborative Certification and Qualification Organisation (CQO) was established in accordance with the contract. This organisation is responsible for establishing airworthiness and qualification requirements, making judgments on their compliance, providing technical advice to OCCAR, and coordinating A400M CQ activities. The structure of the CQO resembles that of EASA (European Union Aviation Safety Agency), with a CQC Chairman, a CQ Committee (CQC) and a team of specialists, including representatives of the Military National Airworthiness Authority (MNAA) and representatives of the OCCAR [46].

The Airbus A400M is often referred to as "dual certified" because it has received two separate certifications for its operation. This dual certification is a significant feature of the aircraft and highlights its versatility and ability to operate in both civil and military contexts. As shown in Figure 2-9, the steps for the certification process of avionics in the A400M consist of:

a) Certification

- <u>Civil certification:</u> the A400M has received civil certification from the EASA or its equivalent in other regions, depending on where it is operated. The first phase focuses on EASA Type Certification, which covers civil type certification, provisions for military systems and EASA compliance. This means that it can be used for civilian purposes, such as commercial cargo and passenger transport, humanitarian missions, and even certain civilian government operations. Civil certification is essential for aircraft to operate in civil airspace and airports.
- <u>Military certification:</u> The second phase involves CQC recognition of the EASA Type Certificate and the military certification process. This phase combines civil certification with military requirements, defining the complete aircraft configuration and specifying military operating conditions and configurations. This includes features such as the ability to carry military cargo, conduct tactical airlift operations, perform aerial refuelling, and meet military airworthiness standards.

b) Qualification

The third phase manages the qualification process, where qualification-related compliance activities are carried out in conjunction with the certification process. Certification compliance testing is used to support qualification testing, ensuring a comprehensive and parallel approach to both civil and military certification and qualification.

While this collaborative approach offers numerous advantages, such as the use of recognised airworthiness codes and established processes, it also presents challenges. Drawing the line between civil and military aspects can be complex, especially in intricate legal environments. In addition, the unconventional design features of the A400M have resulted in a significant number of military certification review items (MCRIs). Nations participating in the program may have to address conflicts with their respective organisations and regulations related to the A400M program.

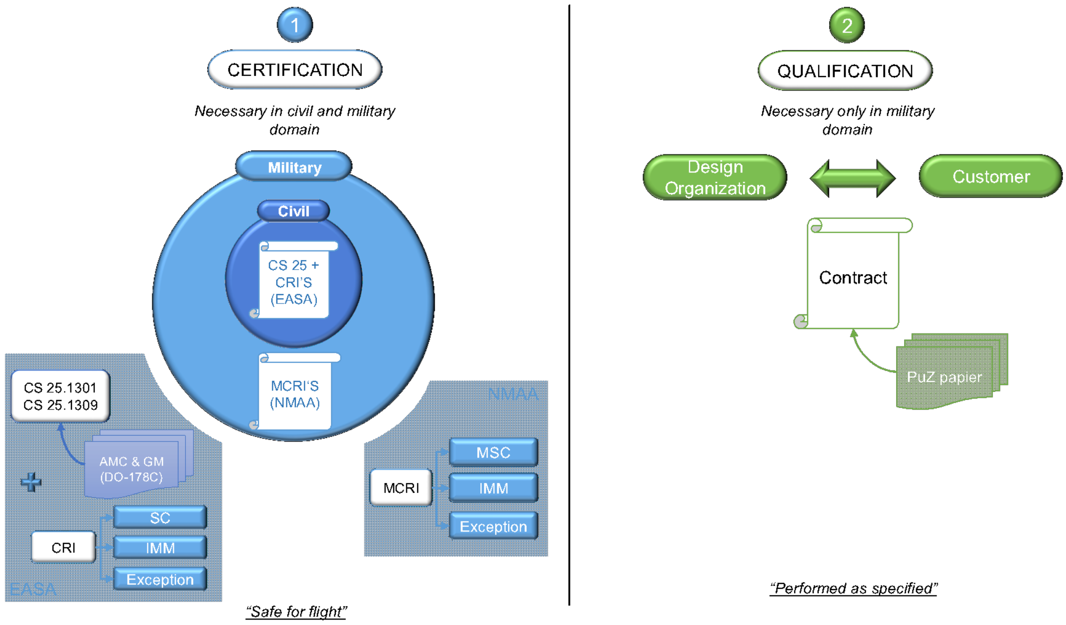

Figure 2-9: Certification basis for airborne software of the military transport aircraft A400M (Source: own elaboration).

2.3 Challenges of certifying AI/ML-based systems in military aviation

The integration of AI applications, concretely ML (sub-)systems, into aviation represents a paradigm shift, disrupting the traditional sequence of software and hardware development. Unlike conventional methods that rely on fixed rules, ML utilises training data to automate behaviour generation and rule extraction. However, this innovative approach poses intricate challenges when certifying ML applications for safety critical aviation functions. In this chapter, we explore the complexities of certifying ML applications within some aviation's concept papers [10], [14], [47], [48]:

- Lack explainability: Explainability refers to the degree to which humans can understand the causality of AI outcomes. Address the limits of predictability and explainability of AI application behaviour, considering the inference mechanisms and model complexity.

- Lack of guarantee of no 'unintended function': develop relevant guarantees on the stability and robustness of AI models and on the absence of "unintended behaviours" in AI applications. Unlike traditional certified software, where the intended function is explicitly defined in the code, the intended function of the ML is embedded in parameter sets obtained through training. Difficulties in keeping a comprehensive description of the intended behaviour and creating a framework for knowledge and data management.

- Paradigm shift between design assurance and learning assurance. Design assurance, essential in certified airborne applications, is intended to demonstrate compliance with safety requirements by evaluating the development process. Learning assurance, specific to ML, focuses on identifying and rectifying errors in data-driven learning to meet applicable requirements. It places importance on defining performance metrics, data management to track requirements, and the choice of model architectures and algorithms.

- Dealing with adaptive learning processes: The integration of real-time learning during system operation introduces significant challenges in ensuring software performance. Existing certification processes are not suited for managing this dynamic behaviour and would necessitate regulatory and guidance material revisions to adapt to it.

- Managing shared operational authority in novel types of human-AI teaming (HAT).

- Addressing model bias and variance trade-off in the various steps of the AI assurance processes.

- Embedding AI models in safety critical systems. This challenge is a consequence of all the previous ones. With all the above challenges, the integration of these ML components into safety-critical systems cannot guarantee the level of safety required by these systems.

- Designing and certifying electronics for these high demanded performance computing applications, for instance, providing certifiable safety critical GPU platforms to allow running AI algorithms efficiently.

- Staff competency, support to the industry and research.

By navigating these challenges, EASA AI Roadmaps and EASA AI Concept papers [4], [10], [13], [48] outlines four crucial building blocks essential for establishing a framework to ensure trustworthy AI in aviation. They provide a comprehensive framework for assessing AI trustworthiness, ensuring safety, providing operational explainability, managing human factors, and mitigating risks associated with AI's complexity.

Military aviation presents a distinct landscape, differing significantly from civil aviation. This divergence shapes the development of software for military applications, driven by three characteristics aligned with military requisites [49]:

I. Mission Accomplishment: A central aspect of military applications is the absolute focus on mission success. Software development is closely linked to the achievement of precise operational objectives.
II. Embracing Innovation: The dynamic field of military aviation demands a keen awareness of emerging technologies and evolving trends. Adapting to new paradigms is an inherent aspect.
III. Secured Content Handling: Military operations are characterised by sensitivity, which requires prudent handling of classified information in various military entities around the world.

While researchers have made giant steps to address the issues of transparency, interpretability and explainability of AI, these advances are applicable to military AI projects. However, it is critical to recognise that military prerequisites may differ markedly and encompass elements such as risk, data integrity, and legal obligations. The applicability of certain transparency modalities may be context dependent.

Certification of AI/ML systems in military aviation may indeed require a different approach compared to civil aviation due to the unique operational and regulatory considerations involved. In this context, the integration of AI into military domains introduces a spectrum of challenges that merit deliberate attention:

- Safety and security of models and data: Military AI systems must prioritise the security of AI models and data used for training and operation. Protection against cyber threats, espionage, and unauthorised access is crucial to maintaining operational security. Military aviation operates in dynamic and adversarial environments where the systems must contend with intentional interference, electronic warfare, and cyber threats.

- Complex environments and difficulty of generalisation: Military operations can involve complex and dynamic environments, including adversarial scenarios in which AI systems may face deliberate attempts to deceive or disrupt their functionality.

- Export control compliance: Military AI systems may include sensitive technologies subject to strict export control regulations. Ensuring compliance with these regulations while collaborating with international partners can be challenging.

- Interoperability and integration: Ensuring compatibility and effective integration is a challenge, especially in multi-platform and multi-vendor environments.

- Data availability and quality: Military applications may require access to high-quality, real-time data, which may be limited in certain operational scenarios. Ensuring data availability and reliability is essential for AI systems to operate effectively.

- Transfer learning: transfer learning makes it possible to adapt pre-trained models to military applications where there is both limited training data and computational resources.

- Human-machine teamwork in high-stress environments: Military aviation often faces high-stress, time-critical situations. Designing AI systems that can collaborate effectively with human operators under these conditions is a unique challenge. The role of humans in military aviation differs from that in civil aviation. In civil aviation, there is typically a clear separation of roles, with pilots being responsible for decision-making and control. In military aviation, AI/ML systems may augment or automate decision-making processes, affecting the certification requirements. Examining the human-machine interaction aspects and defining the roles and responsibilities of both humans and AI/ML systems is vital.

- Model bias and fairness: Ensuring that AI models used in military applications are free of bias and adhere to principles of fairness is crucial to maintaining ethical and legal standards. These considerations may include issues such as accountability, transparency, responsibility, compliance with international humanitarian law, and adherence to ethical guidelines.

- Classification and Categorisation**:** AI/ML systems used in military aviation can vary significantly in terms of complexity, autonomy, and criticality. Developing a framework for classifying and categorising these systems based on their intended use, level of autonomy, and impact on flight safety is crucial for determining appropriate certification processes.

- Public opinion and ethical concerns: the use of AI in military applications often raises ethical concerns and public scepticism. Generating public confidence and addressing concerns about the role of AI in warfare is a major challenge.

- Rapid technological advances: AI and technology in military aviation are rapidly evolving fields. Keeping up with advances against enemies, both in terms of technology and potential threats, is a constant challenge.

- Training and skills development: Military personnel need training to effectively operate and maintain AI-based systems. Ensuring a sufficient pool of qualified personnel is a critical challenge.

- International collaboration and regulation: International collaboration and adherence to evolving regulations and treaties related to military autonomous and AI-based systems require careful consideration and coordination. Military organisations often have separate regulatory bodies, such as defence ministries or military aviation authorities, which establish their own standards and certification processes. Identifying these regulatory materials and understanding any specific requirements is essential in assessing the certification process for AI/ML systems in military aviation.

3 State of Science

The current regulatory framework for hardware and software does not adequately address the unique requirements of AI-based systems. As a result, there are ongoing efforts to update and enhance the existing standards. This chapter provides an overview of the latest advancements in AI technology within the context of avionics and an outline of the existing and ongoing regulatory framework for certifying AI-based systems in the context of civil and military domains. In general, efforts in the civil domain precede military actions.

3.1 AI/ML technology readiness in aviation

Numerous cutting-edge AI avionics projects are currently in development, undertaken by various companies, organisations, and research associations, with the aim of showcasing the vast potential of this technology. It is important to emphasise that without technology readiness and extensive research and development efforts, standardising the implementation of these technologies would be challenging.

In August 2020, an AI-controlled F-16 designed by DeepMind, a Google subsidiary, achieved a remarkable victory over a human F-16 pilot in an aerial dogfight conducted in a virtual reality simulation [50]. The AI-controlled aircraft, trained through online reinforcement learning, scored a decisive 5-0 victory. The dogfight was part of DARPA's Alpha Dogfight competition, which is designed to advance the agency's efforts to build trust in AI and develop manned-unmanned teaming capabilities.

Another example is the ACAS X, which represents the Next-Generation Airborne Collision Avoidance System. It issues horizontal turn recommendations to prevent potential collisions with other aircraft. This system encompasses various versions, including ACAS Xa designed for large aircraft, ACAS Xo, ACAS Xu tailored for unmanned aircraft, and ACAS sXu specifically intended for small, unmanned aircraft. In contrast to the Minimum Operational Performance Standards (MOPS) set forth by EUROCAE in 2020, which advocate for an implementation relying on lookup tables (LUT) without any machine learning components, recent research has demonstrated the efficacy of NNs in addressing the collision avoidance challenge. The primary goal is to reduce the size of the embedded code while enhancing the anti-collision performance [51], [52]. ACAS X is one of the case studies included in the ongoing EUROCAE WG-114/ SAE G-34 certification efforts.

To continue, Daedalean is at the forefront of advancing technologies that drive autonomy, particularly at the intersection of aerospace and AI. Their primary focus is on reengineering the electronic systems in aircraft to enable AI-driven intelligence within the cockpit. Daedalean presents a certifiable embedded system that processes data from high-resolution cameras and various sensors using machine learning, such as detecting traffic in airspace. They discuss the need for such equipment with a real-world example of a machine-learned visual awareness system. Daedalean visual awareness suite accounts for 4 different tasks: visual traffic detection, visual landing guidance, visual positioning, vertical visual landing guidance [53], [54].

In military aviation, the A400M Missile Warner embeds a NN called Multi-Colour Infrared Alerting Sensor (MIRAS) [55]which classifies the infrared images, provided by dedicated IR cameras, as missile threats. This NN has been trained with supervised learning using recorded image information with overlaid missile signals. The equipment (including both software and complex electronic hardware) has been certified in accordance with the applicable certification basis (i.e., CS-25 intended function and safety requirements). In particular, the software has been assessed against DO-178, with no specific consideration on data-driven. On top of the certification exercise, an additional demonstration activity has been performed against the equipment performance requirements specified in the contract. This additional exercise has involved further training of the NN with respect to dedicated missile types (as specified in the contract).

3.2 AI/ML regulatory framework

3.2.1 AI/ML regulatory framework in civil domain

Existing regulatory frameworks for hardware and software items fail to provide adequate acceptable means of compliance for AI-based systems. Therefore, to address this and to facilitate the effective adoption of AI technology, the EU published the EU AI Act [11] as a regulatory proposal that aims to create a harmonised framework for the development and use of AI across the EU. In the aviation industry, EASA and the FAA are working on concept papers [4], [21] to establish initial objectives for anticipating future guidance and requirements related to safety-related ML applications. They are also collaborating with international technical committees like EUROCAE WG-114 and SAE G-34 [2] to develop new industry standards to assist with the development and certification of aeronautical systems using ML technology. As this process may take several years, for the immediate implementation of the ML there are other approaches that contemplate the certification of avionic systems with the conventional standards. For example, the Flight System Dynamics (FSD) institute at the Technical University of Munich developed a workflow to certify ML systems for low criticality avionics applications with the existing standards [56]–[58]. In the framework of recent certification projects, EASA has received proposals to embed an NN trained through DL techniques as part of their system development. Thus, EASA issued a special condition SC-AI-01[59] as an interim solution for certifying specific AI systems in small aircraft. The development of research projects such as MLEAP [60] and the support to industry through Innovation Partnership Contracts (IPCs) and Memoranda of Cooperation/Understandings (MoCs/MoUs) such as Concepts of Design Assurance for Neural Networks (CoDANN and CoDANN II) [5], [20] or DEEL [14] are different examples of tools that EASA uses to collaborate on innovation.

3.2.1.1 EU AI Strategy: EU AI Act

In June 2018, the European Commission established the High-Level Expert Group on Artificial Intelligence (AI HLEG) with the main objective of supporting the implementation of the European AI Strategy regarding the ethical approach to AI. The group's tasks included providing insights on potential policy measures, ethical considerations, legal aspects, and broader societal implications related to AI, including its socio-economic dimensions. In March 2019, the AI HLEG presented a set of seven fundamental principles to ensure the trustworthy AI. These principles are detailed in a report entitled "Ethical Guidelines on Artificial Intelligence" [61]. Afterwards, AI HLEG presented a practical tool consisting of 131 questions to assess the trustworthiness of an AI system under development. This initial assessment checklist was subsequently expanded with the "Final Assessment Checklist for Trustworthy Artificial Intelligence", which was made available in July 2020, along with a web tool called " Assessment Checklist for Trustworthy Artificial Intelligence (ALTAI)" [62].

In the following Figure 3-1, it is possible to see the adoption of these EU ethical guidelines by the aviation sector through the EASA AI Trustworthiness Block.

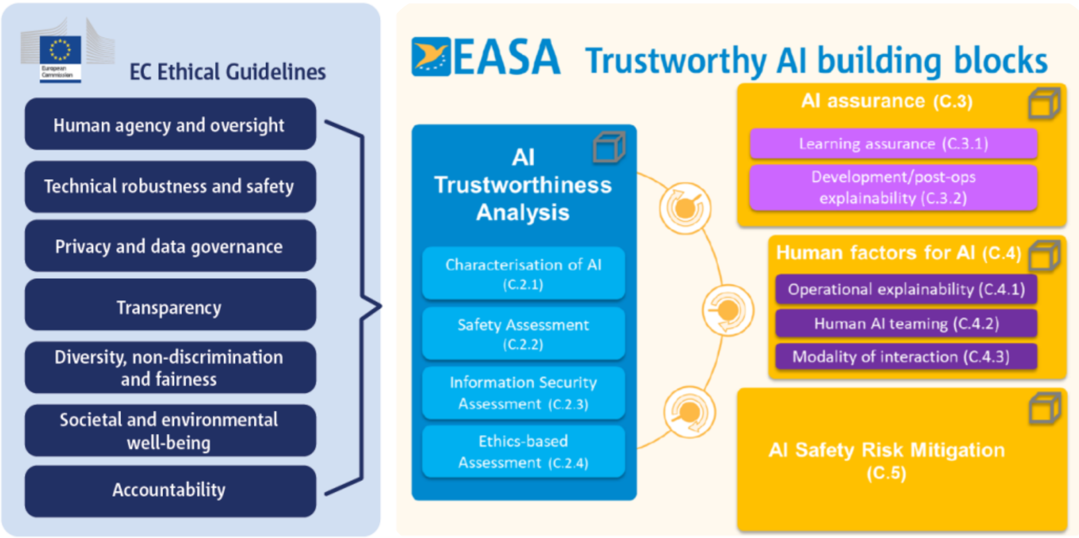

Figure 3-1: EASA trustworthy AI building block and EU AI Ethical guidelines alignment [10]

In April 2021, the EU Commission published the “Proposal for a Regulation of the European Parliament and of the Council on harmonised rules for artificial intelligence “(EU AI Act) [11] as a regulatory proposal to establish a standardised framework for the development and deployment of AI within its member states. The EU AI Act proposes a comprehensive framework for AI governance, encompassing risk assessment, unacceptable AI practices, transparency, data quality, and human oversight. This proposed regulation has the potential to reshape AI practices across the EU. The European Parliament adopted its negotiating position on the AI Act with 499 votes in favour, 28 against and 93 abstentions ahead of talks with EU member states on the final shape of the law [63]. Regarding aviation regulations, Article 81 of the EU AI Act anticipates the incorporation of its requirements into future acts related to airworthiness, ATM/ANS, and unmanned aircraft. EASA AI Concept Papers [4], [48] align with this act to ensure compatibility within EASA's domains.

3.2.1.2 EASA papers on AI

EASA has published several important documents regarding the certification of ML-based systems. These include the "EASA AI Roadmap 1.0" [13] in February 2020, aimed to identify opportunities and challenges brought by AI in aviation across various domains, as well as guidance documents entitled "First usable guidance for Level 1 machine learning applications" [48] in December 2021 and "First usable guidance for Level 1 & 2 machine learning applications" [4] in February 2023. In addition, EASA introduced the SC "SC-AI-01 Trustworthiness of Machine Learning based Systems" [59], which mandates the use of the "First usable guidance for Level 1 machine learning applications". And its most recent publication, the "EASA AI Roadmap 2.0” [10], was issued in February 2023 in advance of the required update due to advances in AI technology and evolving perspectives from the aviation industry.

In essence, these materials outline EASA's approach to addressing the challenges of certifying applications that rely on ML. This includes establishing clear guidelines for different levels of ML applications and addressing the unique considerations and risks associated with these systems in the aviation context [47].

➢ ***Roadmap 1.0 and Roadmap 2.0***

The “EASA AI Roadmap 1.0” [13] has been updated due to the rapidly advancing technology in AI and the evolving perspectives of the aviation industry. Refer to Figure 3-2 for the last update of EASA AI Roadmap.

For version 2.0 of the AI Roadmap [10], EASA, in line with Annex I of the proposed EU AI Act, classifies AI techniques into three categories described in more detail previously in section 2.1.1. While the first version of the AI Roadmap emphasised machine learning, particularly deep learning, the update recognises that stringent aviation safety regulations are motivating certain stakeholders to explore new knowledge-based AI solutions. This change serves as a primary motivation for EASA to revise this document. Importantly, it's worth noting that these different AI approaches can be combined in hybrid AI configurations, which also fall within the scope of this roadmap. As a result, the roadmap has been extended to include all the techniques and approaches (refer to Figure 2-2).

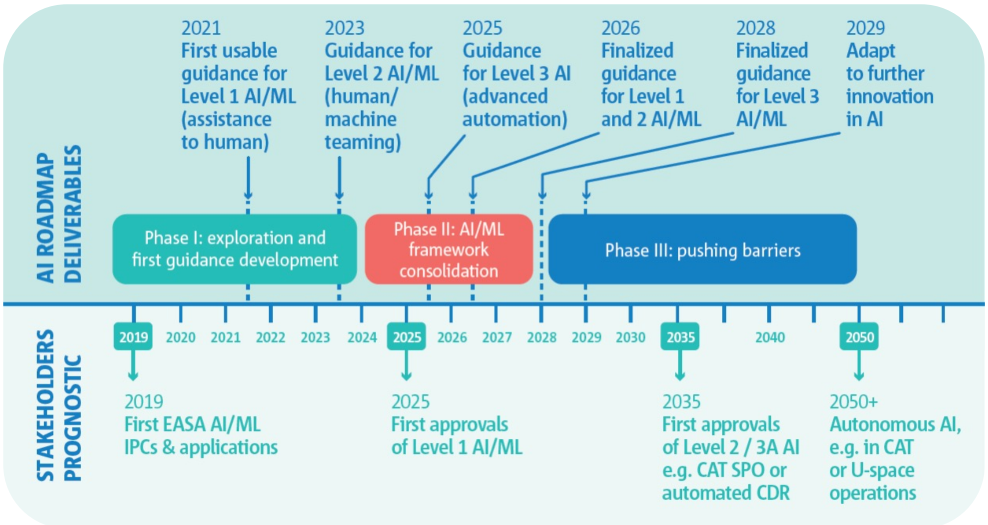

Figure 3-2: Last version of EASA AI Roadmap [10]

➢ ***AI Concept papers***

EASA AI Concept papers [4], [48] outline four crucial building blocks (indicated Figure 3-1) essential for establishing a framework to ensure trustworthy AI in aviation and to facilitate the safe and effective utilisation of AI/ML technologies within the industry.

I. **AI Trustworthiness Analysis Building Block:** This initial step involves characterising the AI application, conducting ethics-based assessments, safety evaluations, and security analyses. These assessments are fundamental prerequisites for any AI/ML system development and are integral processes leading to the approval of innovative

AI solutions. The building block connects with EU Ethical Guidelines and serves as a gateway to the subsequent technical building blocks.

- Characterisation of AI
- Safety Assessment
- Information Security Assessment
- Ethics-based Assessment

II. **AI Assurance Building Block:** This block focuses on AI-specific guidance for the system. It includes two key aspects: learning assurance, addressing the shift from traditional programming to AI learning processes (see **Error! Reference source not found.** for the W-shaped process), and development/post-ops explainability, ensuring the ability to provide understandable, reliable, and relevant information about how the AI system produces its results. Additionally, data recording capabilities are emphasised for continuous safety monitoring and incident/accident investigation support.

- Learning assurance
- Development/post-ops explainability

III. **Human Factors for AI Building Block:** This block introduces guidance to address human factors specific to AI integration. It emphasises AI operational explainability to provide human users with understandable information about AI/ML outcomes. The concept of human-AI teaming is also introduced to ensure effective cooperation and collaboration between human users and AI systems for achieving desired goals.

- Operational explainability
- Human AI teaming
- Modality of interaction

IV. **AI Safety Risk Mitigation Building Block:** Recognising the potential limitations in fully understanding AI processes, this block focuses on addressing residual risks associated with AI's inherent uncertainty. It acknowledges that complete transparency might not always be achievable, and strategies are needed to manage the residual risks effectively.

➢ ***SC-AI-01 "Trustworthiness of Machine Learning based Systems"***

As referenced in [47], a Special Condition (SC) is established by EASA when applicable airworthiness regulations do not comprehensively address safety requirements due to novel or unconventional design features.

In the area of ML applications, EASA has introduced the Special Condition referred to as "SC-AI-01 Trustworthiness of Machine Learning based Systems" [59]. This guideline aims to assist manufacturers, developers and operators in identifying and mitigating risks associated with ML-based systems. To achieve this goal, SC-AI-01 specifies the use of the "Initial guidance for Level 1 machine learning applications". This guidance is intended to supplement the CS

23.2500 & CS 23.2510 for applications incorporating AI/ML technology. The CS 23 [64] covers the certification requirements for aircraft in the normal, utility, aerobatic and commuter categories. As such, it has equivalent status to other related material within the EASA regulatory framework. These specific sections have been found lacking in providing adequate criteria for ML-based systems, particularly about issues of AI interpretability and ethics-driven evaluations. The SC-AI-01 can be seen as an interim point of reference until dedicated regulatory material on AI is formally established.

Some aviation companies, such as Daedalean have already applied for this SC and are in the middle of a type certificate, as stated by its CEO in the webinar “EASA AI Days High Level Conference” in May 2023 [65].

3.2.1.3 EUROCAE/SAE ARP6983

EUROCAE WG-114, in collaboration with SAE G-34, was established with the primary objective of facilitating the safe and effective integration of AI technologies into aeronautical systems. The joint international committee WG-114/ G-34 oversees developing the new industrial standard ARP6983 “Process Standard for Development and Certification of Aeronautical Safety-Related Products Implementing Artificial Intelligence” [66]. The working group is focused on evaluating critical applications of AI within aeronautical systems, particularly those embedded in aircraft and used in ground equipment. Its goal is to create standards that support the development of safe, compliant systems.

The joint group has grown rapidly around the world and now includes more than 500 engineers. It acts as a bridge between industry stakeholders and regulators to create consensus standards that can be accepted by certification authorities and adopted by the industry at large. The group draws on expertise from a wide range of sectors, including major aircraft manufacturers, unmanned aircraft systems, urban air mobility and electric vertical take-off and landing systems, engine manufacturers, as well as regulators and air navigation service providers. Prominent names such as Airbus, Boeing, EASA, FAA, NASA, DoD, Eurocontrol, Thales, Dassault, Safran, Nvidia and Intel are active participants, fostering collaboration between large corporations, smaller companies, and institutions. The organisational structure of the working group is carefully designed to ensure balanced representation, with leadership and executive roles filled by both avionics and ground experts.

In 2021, the working groups introduced the AIR6988/ER-022 “Artificial Intelligence in Aeronautical Systems: Statement of concerns” [2] to address common concerns across the aviation industry and highlight the key challenges that currently hinder the use of AI and machine learning in safety critical applications. To continue, the publication of the AIR6987 "Artificial Intelligence in Aeronautical Systems: Taxonomy" [67] is planned for the end of 2023 and will provide a comprehensive taxonomy of AI in aviation and facilitate understanding of the concepts (see Figure 3-3). A document with possible AI technologies, architecture, validation approach, and safety concerns for different use cases will follow “Artificial Intelligence in Aeronautical Systems: Use Cases - AIR6994”. To finish with the documents, it is expected to have the standard ARP6983 published by 2024 (see Figure 3-3).

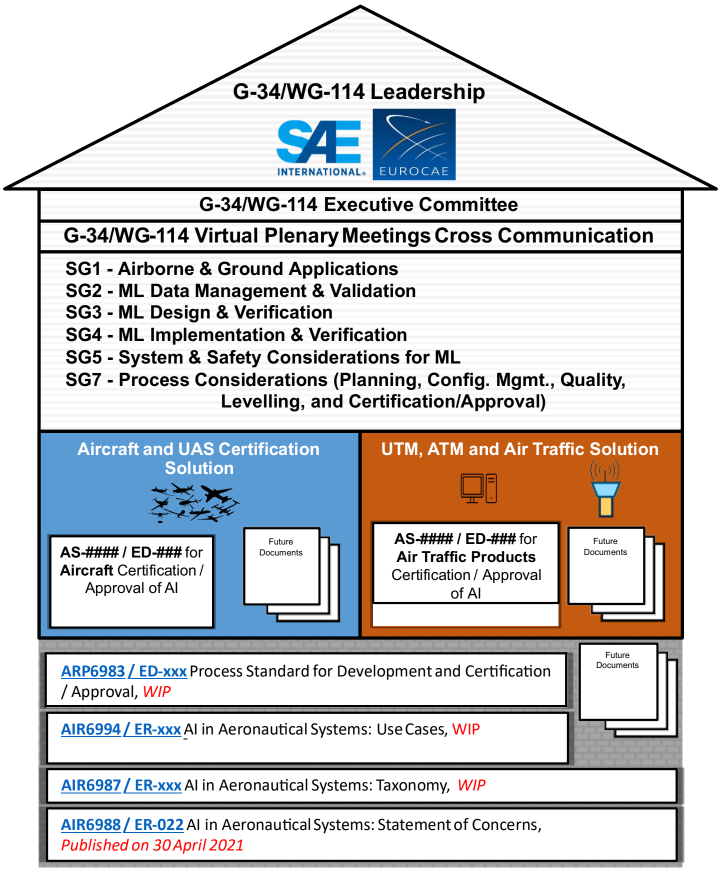

Figure 3-3: WG 114 scope [2], [66], [67].

This standard presents a structured process called the Machine Learning Development Lifecycle (MLDL), like established aviation guidelines such as the ARP4754 or the DO-178C. This MLDL is based on EASA's W-shaped process framework. The MLDL begins with the ‘ML Constituent Operational Design Domain Process’, which defines the scope and context of the AI/ML system. Next, the ‘ML Constituent Requirements Process’ shapes both high-level and derived requirements based on system specifics and design considerations. Next, the ‘Data Management Process’ focuses on producing quality datasets for training, validation, and testing, while also outlining data processing requirements. This is followed by the ‘ML Model Design process’, which aims to iteratively develop the architecture and requirements to be used to build, train, and optimise the ML model. Once the model is formulated, the ‘ML Validation and Verification Processes’ aim to determine whether both the ML data and the ML model comply with the validated ML requirements. Verification includes assessments and testing. With a validated and verified ‘ML Model Description’, the model can be implemented using existing software and hardware development processes. Once the software and hardware components are integrated, the next steps include ‘ML Model Integration’ and, ultimately, integration of the entire AI/ML system into the larger aviation system [47].

While EASA, in its concept papers, proposes high-level objectives for ML development and approval, the WG-114/EG-34 is preparing detailed technical industry standards for ML. The ARP6983 serves as an intermediate engineering layer between the system lifecycle process e.g., ARP4754 and the implementation lifecycle processes e.g., DO-178C, DO-254 with a specific focus on machine learning-based systems.

3.2.1.4 Industry (IPC) and academy

- ***CoDAAN I (2020) & CoDAAN II (2021)***

The “Concepts of Assurance for Design of Neural Networks (CoDANN)” reports [5], [20] jointly published by the EASA and Daedalean in March 2020 and May 2021, were the result of two Innovation Partnership Contracts (IPCs), which are research, innovation and industry engagement tools used by EASA. The CoDANN reports looked at how existing guidelines such as [ED-79A/ARP4754A; ED-12C/DO-178C] could be applied to systems incorporating AI, ML and NNs. They also identified gaps in these guidelines and suggested possible ways to address them. A key outcome of these reports was the development of a novel W-shaped development process, as shown in Figure 2-5. This process adapted the traditional V-shaped cycle to better suit machine learning applications. This unique W-shaped process revolves around the concept of learning assurance and is designed to ensure that following it will enable the ability to provide performance guarantees for systems that incorporate a machine learning component.

The publication entitled "Concepts of Design Assurance for Neural Networks (CoDANN)" [5]begins with an overview of existing guidelines and working groups, followed by a comparison between traditional software and ML software. The document presents an aviation use case and its ConOps. It looks at the risks associated with both traditional and ML-based systems and discusses mitigation strategies. The concept of 'learning assurance' for ML-based aviation systems is introduced, with a novel W-shaped development cycle that extends the traditional V-shaped cycle to include verification of the learning process, training, data lifecycle management and verification steps. Additional concepts such as transfer learning and synthesised data are also explained. The document covers performance evaluation through metrics and model evaluation. A critical premise of ML learning is to ensure the accuracy of the input distribution in the training data. Identifying the probability distribution for the intended function is essential to provide functional guarantees, paralleling guidance from aircraft and automotive design documents. Metrics for evaluating learning safety are presented, including the confusion matrix detailing true/false positives and negatives, as well as higher-level metrics such as precision and recall. Finally, the document emphasises the importance of safety assessment to evaluate the risk of systems incorporating ML components.

The EASA CoDANN II report [20], builds on the first report and focuses on extending the concept of learning assurance. Concretely, it focuses on the implementation of models and inference environments. Verification processes must consider the significant differences between the learning environment in which training takes place and the inference environment in which the function is used for its intended purpose. The report introduces the concept of explainability, which is about providing understandable and relevant information about how AI/ML applications produce results. Explainability is divided into system-level and output-level explanations that support learning assurance and operational aspects. The report also addresses security assessment, highlighting the importance of out-of-distribution (OOD) detection. The notion of input data distribution, discussed in the previous report, is formally defined. Identification of the target distribution becomes critical to identify input data during operation that deviates from these assumptions. Ensuring that the system recognises and communicates untrustworthy outputs in such cases is critical. One approach is to use a runtime assurance architecture such as ASTM F3269-17 [68], which is also described in the report.

➢ ***DEEL (2021)***

The DEEL (Dependable Explainable Learning) Project, which engages both academic and industrial collaborators. This initiative is led by IVADO, IRT Saint Exupéry, CRIAQ, ANITI and IID Laval, with industrial partners including Airbus, Thales, Renault, and Continental. The main objective of the DEEL project is to facilitate the development of reliable, robust, explainable, and certifiable AI components for use in critical systems.

In March 2021, the DEEL certification workgroup released a whitepaper titled "Machine Learning in Certified Systems" [14]. This document outlines various methods, concerns, and potential solutions related to the application of machine learning in certified systems.

➢ ***FAA (2021)***

The FAA published the document “Neural Network Based Runway Landing Guidance for General Aviation Autoland” in 2021 [21]. This document outlines a collaborative research project conducted by Daedalean AG and the FAA between September 2020 and September 2021. The project aimed to develop and test a visual landing system (VLS) for fixed-wing aircraft using AI, ML, and NNs. The research objectives included evaluating the performance of the technology for runway landing guidance in general aviation aircraft, testing compliance with FAA certification processes, informing policy development for machine learning-based aviation systems, validating the system's potential as a backup during GPS disruptions, and contributing to future certification requirements. The research findings guide the FAA in shaping certification policy for AI and NN applications in general aviation, with a focus on reliability, robustness, and real-world suitability.

➢ ***MLEAP (2023)***

The MLEAP project is part of EASA's efforts and is funded under the Horizon Europe framework. MLEAP is designed to address the challenging objectives outlined in the W-shaped process mentioned in the EASA AI Concept Paper.

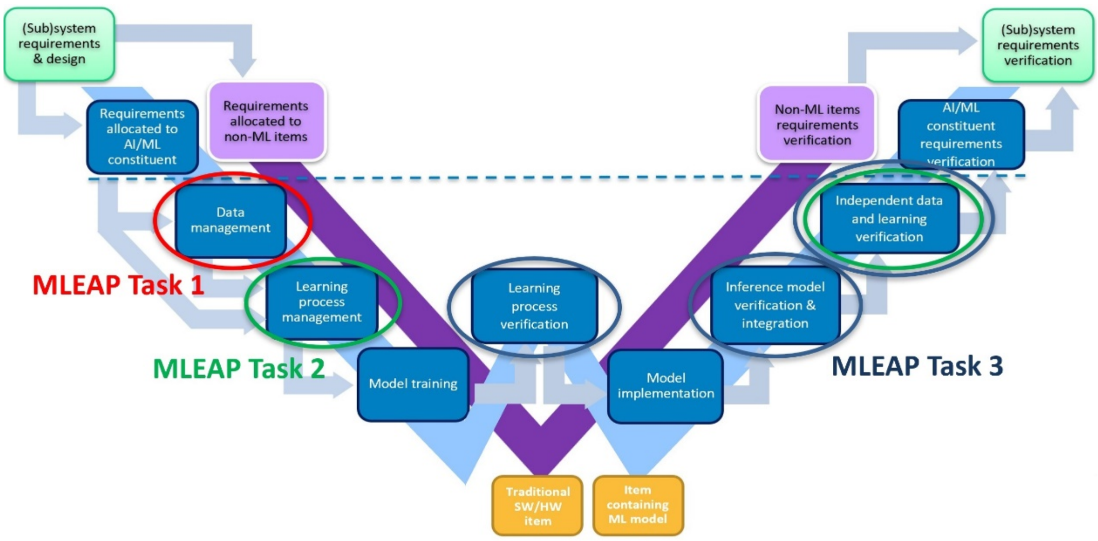

Figure 3-4: MLEAP project three main tasks focused on the W-shaped development process [60].

The MLEAP project consists of three main tasks (see Figure 3-4):

- Task 1: Addressing data completeness and representativeness, including the handling of corner cases.
- Task 2: Developing AI models considering generalisation properties.
- Task 3: Evaluation of AI models, with a focus on robustness and stability.

This interim report “EASA Research – Machine Learning Application Approval (MLEAP)” [60] provides an overview of the expected approaches to the evaluation and certification of AI-based systems in line with the objectives set out in the EASA roadmap. It aims to assist industry stakeholders in developing strategies for the safe use of AI in both its operational and technical aspects within the aviation sector. This publication represents the first public version of the project, with a final and comprehensive edition to be published by May 2024.

➢ ***TUM Flight Dynamics Systems Institute (2021 & 2022 & 2023)***

In the first paper "Toward Certification of Machine-Learning Systems for Low Criticality Airborne Applications" [56] published in 2021 by the FSD Institute analysed the compatibility between existing industry standards for airborne software and the unique challenges posed by low-critical ML systems with DAL D. The study identified potential obstacles related to traceability, coverage, and ML model verification. To address these issues, the researchers proposed a tailored ML development workflow specifically designed for non-adaptive, supervised learning systems. This approach aligns with the established model-based design methodology. The paper concludes that, under the proposed assumptions, all relevant objectives of key standards, including DO-178C, DO-331, DO-254, and ARP-4754, can be achieved for Level D ML-based systems using conventional activities and methods.

In 2022, the institute presented a case study centred on the certification aspects of ML systems in safety critical airborne applications through the paper "Toward Design Assurance of Machine-Learning Airborne Systems" [57]. They introduced the Runway Sign Classifier (RSC), a realistic vision-based application using a deep neural network (DNN) to detect and recognise airport runway signs. The RSC was categorised into DAL D, DAL C, or DAL B based on different use cases. Through this case study, the researchers demonstrated that, by adopting a custom ML development workflow, all necessary objectives for DAL D certification can be met. Additionally, they identified alternative mitigation strategies for DAL C and higher assurance levels to address ML technology's incompatibilities with current design assurance practices.

The latest publication early this year 2023, "Towards Design Assurance Level C for Machine-Learning Airborne Applications” [58], proposed an architectural mitigation technique enhanced by alternative verification methods and development independence to extend the custom workflow for ML-based systems. While reducing the design assurance level using solely architectural mitigation is not recommended due to the novelty of ML technology in safety critical contexts, the researchers conducted a thorough analysis of ML traceability, coverage, and verification issues. They also introduced alternative methods to address problematic assurance objectives affected by these issues. This approach aligns with industry standards that emphasise the combination of redundant practices to ensure safety.

The institute intends to study tool qualification issues and expand the analysis of alternative verification means for higher assurance levels, specifically levels B and A.

3.2.1.5 Rulemaking concept for AI

The Figure 3-5, taken from [47], summarises all the current certification materials being developed (in red) and how they relate to the conventional ones.

Currently under review by the European Parliament and Council, the EU AI Act is expected to be approved as a European binding law and to form the basis for future CSs regulating AI-based aviation systems. To continue, CSs, in conjunction with AMCs and GMs, establish compliance with European civil aviation regulations, specifically EC 1139/2018 and EC 748/2012. These regulations align with the ICAO Standards and SARPs. The SC-AI-01 represents a supplement to CS 23 and specifies the use of the concept paper "First usable guidance for Level 1 & 2 machine learning applications", which serves as AMCs and GMs.

The ARP6983 standard is designed to fit seamlessly into the broader spectrum of industrial standards. It is expected that ARP6983, which operates at the system/sub-system level, will find its place between system-level standards such ARP4754 and ARP4761, and item-level standards such as DO-178C and DO-254. This strategic positioning of the new AI standard ensures its harmonious coexistence with existing standards, without disrupting conventional developments that do not include AI technology.

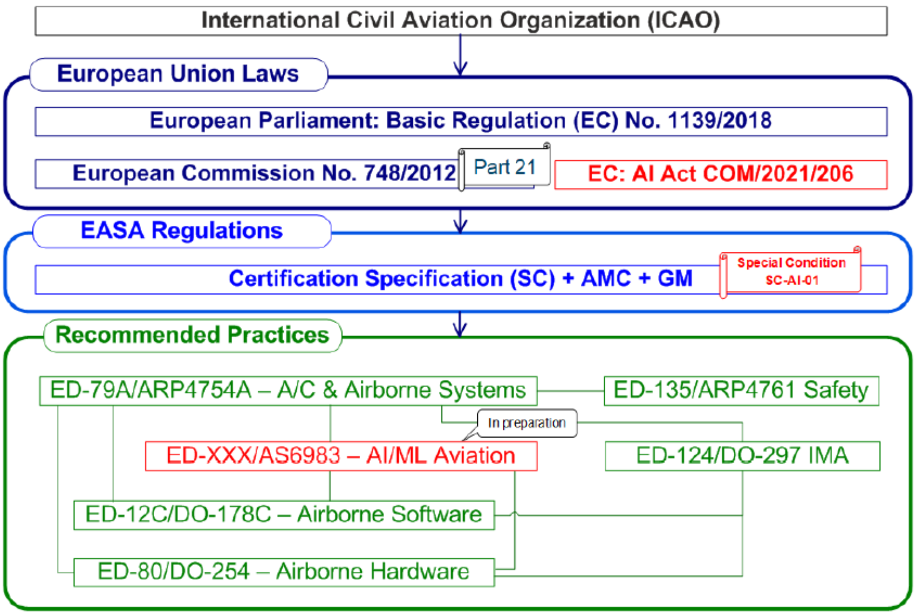

Figure 3-5: AI regulatory material [47].

Having looked at the current regulatory framework, Figure 3-6 shows the desired future regulatory framework. The EASA AI Roadmap 2.0 [10] introduces a two-step regulatory approach, including cross-domain rules (horizontal) and domain-specific rules (vertical).

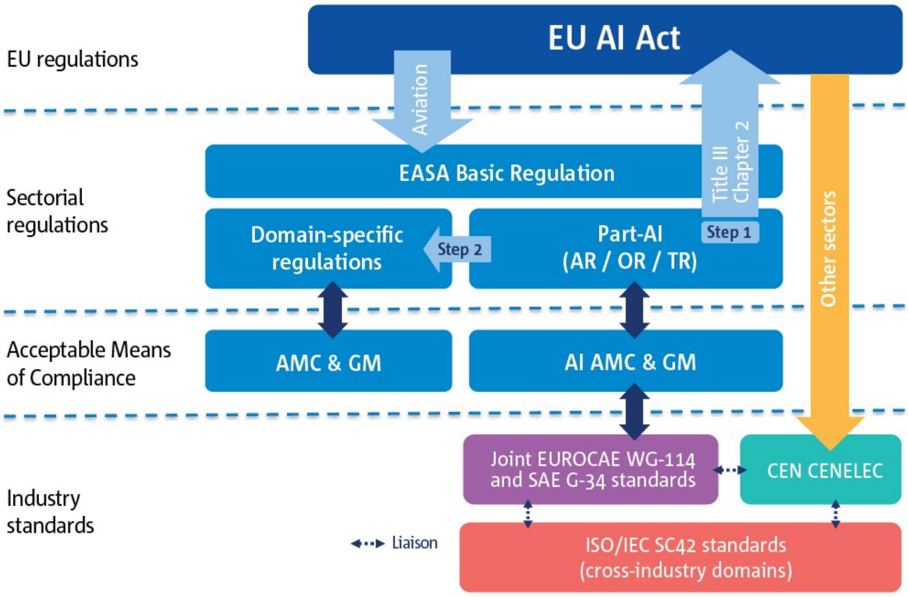

Figure 3-6: Anticipated regulatory structure for AI [10].

This approach consists of:

- **Step 1:** A comprehensive framework, Part-AI, is envisioned, covering regulatory authority (Part-AI.AR), organisational requirements (Part-AI.OR), and AI trustworthiness criteria (Part-AI.TR) from the EASA concept paper. This will include acceptable means compliance (AMC) and guidance material (GM), incorporating the "anticipated MOC" already captured in the EASA AI Concept Papers.
- **Step 2:** Domains are individually assessed for specific and additional requirements. This will ensure a robust basis for the implementation of the new Part AI.

This strategy will also be aligned with the EU AI Act, Title III, Chapter 2 [11].

3.2.2 AI/ML regulatory framework in military domain

Especially in military aviation, there has been a tendency to adopt guidelines and procedures that have been used effectively in civil aviation, such as DO-178C. However, a military project is specified and certified through contracts with mission characteristics and common agreements between stakeholders. This is where the certification process starts to diverge and where gaps, advantages and challenges can be identified in terms of AI certification. As mentioned in previous chapters, in Europe, military and state aircraft are excluded from the scope of the EASA legislation. Consequently, military airworthiness is only regulated at the national level. The need to improve European defence capabilities influenced the decision of the EDA to create the MAWA forum to develop, adopt and implement harmonised EMARs. Regarding the implementation of ML-based systems, MAWA forum has not yet started to address the issue, as its work is mainly focused on the adoption of civil aviation legislation which is still under development. On the other hand, NATO Advisory Group (NIAG) has already initiated the development of a responsible and user-friendly AI certification standard (NIAG-SG279) [69] to help industries and institutions across the Alliance ensure that new AI and data projects are compliant with international law, as well as NATO standards and values in terms of governance, traceability, and reliability. In the field of AI implementation and standardisation in the military sector, the EU has provided funding and co-funding for various projects, including initiatives such as the European Initiative for Collaborative Air Combat Standardisation (EICACS) [70] and AI for Defence consortium (AI4DEF) [71].

3.2.2.1 EDA & MAWA

Although MAWA forum has not initiated efforts to tackle the implementation of AI/ML-based avionics, EDA has already taken a leading role in advancing the field of general military missions (both air, land, and sea) through the Defence Applications of AI (DAAI) Workshop [72]. Through its annual DAAI Workshops, the EDA provides a significant platform for experts, researchers, and practitioners to explore the latest advancements in AI relevant to defence and security. The most recent iteration, the 4th DAAI Workshop “AI Action Plan Workshop: Trustworthiness and Standardisation for AI in Defence” [73], which took place in September 2023, highlighted recent advancements in AI relevant to defence and security. This year's focus was on practical AI applications and scenarios that address decision-making, situational awareness, and other pertinent challenges in military operations. Renowned industry and academia speakers shared AI solutions that have demonstrated significant value and effectiveness in diverse military contexts. As AI's transformative potential for defence and security becomes increasingly evident, this workshop aims to assess the current state of applicable AI solutions, enabling a better understanding of future requirements and capabilities. As defence and security environments become more intricate and data-intensive, the workshops aim to foster collaboration and insights to ensure the effective integration of AI technologies. The EDA's commitment to these workshops underscores its dedication to promoting innovation and shaping the future of defence and security domains through AI.

3.2.2.2 NATO & NIAG-SG279

In the NATO AI Strategy [6], it is stated that all the allies that want to implement AI technology into their military systems, must comply with the six AI Principles of Responsible Use (PRUs):

I. **Lawfulness:** AI applications shall adhere to both national and international legal frameworks, including pertinent human rights and humanitarian laws.

II. **Responsibility and Accountability:** The development and utilisation of AI applications shall involve prudent judgment and care, with clear human oversight to ensure accountability.

III. **Explainability and traceability:** AI applications should possess appropriate clarity and transparency, incorporating review methodologies, sources, and procedures. This involves validation and verification mechanisms, potentially at the NATO or national level.

IV. **Reliability:** AI applications must be rigorously tested and assured for specific use cases, encompassing safety, security, and robustness throughout their lifecycle. This should encompass established certification processes, whether at a NATO or national level.

V. **Governability:** AI applications should be aligned with their intended functions, enabling suitable human-machine interaction and the capability to identify and address unintended consequences. Mechanisms for system disengagement or deactivation due to unintended behaviour should also be in place.

VI. **Bias Mitigation:** Proactive measures should be taken to minimise unintended bias in both AI application development and data sets.

The Data and Artificial Intelligence Review Board (DARB) was established by NATO to govern the responsible development and use of AI in accordance with NATO's six PRUs. The NIAG-SG279 [69] has gathered civil and industry standards, best practices, and recommendations from various experts to assist the DARB in developing a certification process to ensure that the six AI PRUs are met by all military applications used by the Alliance. According to NIAG-SG279, the certification methodology will be tested against six different use cases, providing valuable lessons and recommendations for implementation. However, after speaking with experts involved in the completion of the standard, no safety critical military airborne systems have been reported.

3.2.2.3 FCAS & EICACS

France, Germany, and Spain are collaborating on the Future Combat Air System (FCAS) [25], a system-of-system program that integrates and combines manned and unmanned, existing, and future platforms, and tools. The FCAS program is a constellation of interconnected systems, in particular a Next Generation Fighter (NGF) in collaboration between unmanned and manned systems through an intricately interconnected combat cloud. FCAS represents a powerful, innovative, and fully European joint weapon system solution to ensure combat system superiority. The next-generation system-of-systems architecture will provide a more digitised, collaborative, and agile operating environment. FCAS will use AI and neural networks to enable the aircraft to collaborate with unmanned platforms, share information with other players through a multidomain combat cloud, and assist pilots with situational awareness and decision-making. Customer expectations for the use of AI and autonomy in FCAS mainly revolve around achieving information superiority, faster decision making and streamlining the OODA loop. At the same time, particular emphasis is placed on the importance of maintaining human control and ensuring greater protection of both military personnel and civilians. However, the intersection of these expectations with the debate over lethal autonomous weapons systems (LAWS) raises significant challenges and concerns. This debate encompasses legal aspects related to International Humanitarian Law (IHL) and Human Rights Law (HRL), as well as the concept of meaningful human control over AI-powered systems. It is essential to incorporate means of monitoring and human intervention using a clear and consistent concept of human responsibility. This concept must provide users with a comprehensive understanding of the accountability and ethical responsibility associated with every decision made [25].

The white paper titled "The Responsible Use of Artificial Intelligence in FCAS – An Initial Assessment" [25] aims at contributing to the FCAS Forum debate by providing a technical view on different ethical principles as well as their application on potential use cases through the exemplary employment of the ALTAI methodology (explained previously in EU AI Strategy: EU AI Act chapter).

On the other hand, the EICACS project [70], co-funded by the European Union, seeks to establish Collaborative Air Combat standardisation by ensuring interoperability among heterogeneous assets and seamless collaboration of future air combat systems, encompassing manned, unmanned, and legacy platforms along with sensors and effectors. With a consortium of 11 countries, the project aims to define design rules, interoperability standards, and functional interfaces while addressing challenges posed by AI implementation. The results of this coordination program will likely benefit the two “system of systems” programs currently being developed in Europe, namely the Future Combat Air System (FCAS) regrouping France, Germany, and Spain, and the Global Combat Air Programme (GCAP), the result of a merger between Italy and UK’s Tempest initiative and Japan’s F-X program.

EICACS endeavours to enhance mission execution and performance within systems of systems contexts while promoting the efficient integration of new capabilities into mission system software. The EICACS project aspires to push collaborative air combat to new heights, aligning with European requirements for highly integrated multi platform mission management capabilities.

The general objectives of the EICACS project are twofold:

a) On the one hand, the advancement and establishment of design principles and interoperability standards to ensure a seamless collaboration between assets within the NATO and coalition frameworks. This means facilitating the secure sharing of resources, fostering superior data sharing (combat cloud), and enhancing decision-making capabilities. In addition, the reinforcement of the scalability of the mission system software, the contribution to the interoperability of the mission system execution platform and defining physical and functional interfaces for effectors and sensors, collectively striving to enable cohesive and efficient air combat operations. This requires standardised communication protocols between assets and their components, while accommodating for continuous evolution and rapid integration of new capabilities.

b) On the other hand, the EICACS project undertakes the assessment of questions introduced by the implementation of **AI technologies (enabler)**, which encompass:
   I. **Addressing airworthiness and safety concerns associated with dependable AI-based functions. (scope)**
   II. Promoting European sovereignty over AI engineering tools, algorithm libraries, and methodologies pivotal for leveraging AI in military assets.
   III. Ensuring compatibility of tools and processes throughout the stages of development, validation, qualification, and certification of AI-driven operational services.

EICACS is focused on providing the tools necessary to execute the FCAS program, in which the combat cloud and AI are technology enablers. Most likely, in terms of AI technology standardisation, they will refer to the civilian standard ARP6987.

3.2.2.4 AI4DEF

The AI4DEF consortium [71] is a collaborative initiative of 21 partners from 10 European countries, which aims to showcase the effectiveness and advantages of AI systems in enhancing situational awareness for critical decision-making across various scenarios. Through four distinct use cases, the project will demonstrate both functional capabilities and technological assessments, drawing from existing solutions and military needs. These use cases span multiple defence domains, encompassing air, sea, and land operations. The use case encompassed in the scope of this work (avionics) is the use case 1 which focuses on AI to improve UAV ISR (Unmanned Aerial Vehicle, Intelligence, Surveillance, and Reconnaissance).

Key to the project's success is the AI4DEF LabStore, which plays a crucial role in scaling up AI integration within defence systems. This LabStore is essential for managing growing complexities as technology advances. AI4DEF will facilitate the incorporation of AI-based analytics, enhancing defence capabilities through:

- **Innovative AI Technique Selection**: A unique methodology will be employed to select AI techniques, considering all aspects of the system, and addressing technical challenges. Focus will be on ensuring understandability and trustworthiness.
- **Optimised Human Interactions**: The project will identify optimal ways for human-AI interaction in operational contexts. Ethical, legal, and standardisation criteria will be considered.
- **Comprehensive AI Catalogue**: AI4DEF will offer an extensive range of AI techniques and analytics, ensuring the selection of the most suitable AI defence systems.

3.2.2.5 KIEZ 4.0

The Künstliche Intelligenz Europäisch Zertifizieren (KIEZ 4.0) [74] project focuses on harnessing the potential of AI applications to improve aviation processes while maintaining high safety standards. The project aims to develop AI-based applications for aviation and demonstrate their certification potential. It aims to identify safety, certification, and authorisation mechanisms applicable to AI in aviation, considering both traditional aviation certification processes and those from other industries with AI components. The project involves a consortium of eight partners, including aviation, research, IT, legal and education experts, working together at national and international level. The consortium, in partnership with EASA, aims to take steps towards the implementation of AI certification. Specifically, the Institute of Flight Guidance, in cooperation with Deutsche Flugsicherungs GmbH and using the Air Traffic Management and Operations Simulator (ATMOS), is developing and testing certification concepts for AI applications in air traffic control and demonstrating their benefits.

As a summary and following the increasingly used methodology of 'dual-certification', Figure 3-7 presents a current certification military avionics framework including the efforts in terms of AI-based systems (marked in yellow).

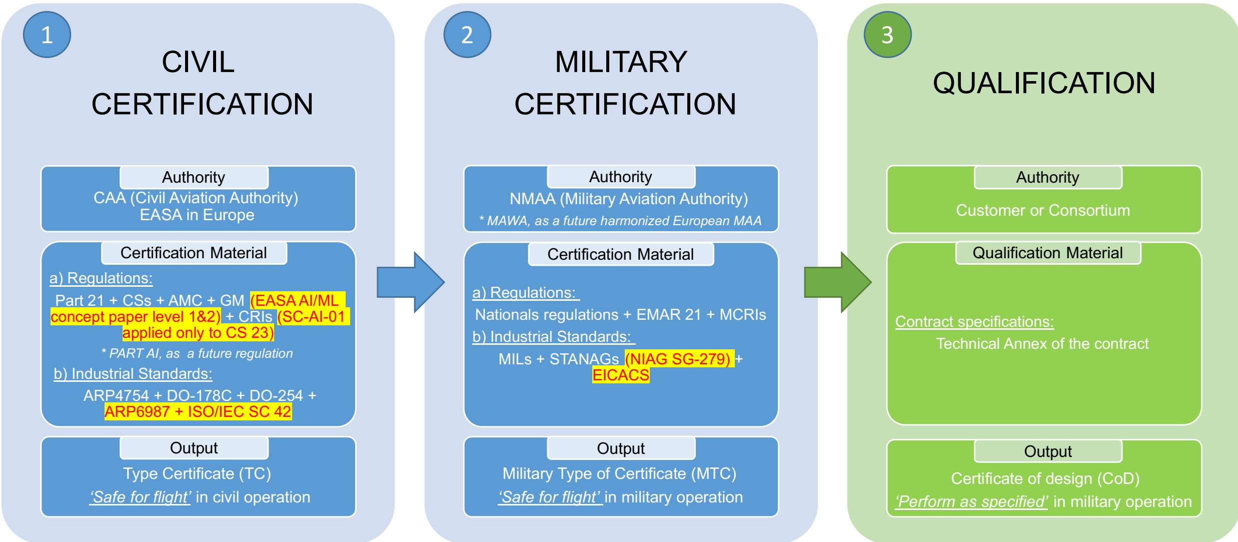

Figure 3-7: On going AI standardisation efforts (Source: Own elaboration).

4 Methodology

To assess the certification framework of a military AI-based system, it is important to first highlight the expected outcomes of this assessment. Once these outcomes have been defined, a certification framework will be presented to integrate, correlate, and derive these outcomes.

The three main expected outputs of this thesis are as follows:

I. Synthesis of current efforts: Synthesis of the existing and upcoming regulatory framework for certifying AI-based systems in the context of civil and military domains.

II. Alignment with the concept of dual certification: The methodology must align perfectly with the concept of adapting principles from the civilian domain to the military context. This alignment is applied not only with the objective of ensuring the applicability of civilian guidelines to military contexts, but a double recognition by both domains. This approach serves to streamline and accelerate the program, resulting in a reduction of the time and cost investment.

III. Proposal of comprehensive recommendations for the certification of ML-based systems in military aviation: A significant contribution of this research is the proposal of ground-breaking guidance material tailored to the military domain, specifically addressing the complex use of AI.

This regulation framework encompasses both horizontal and vertical workflows for implementing a machine learning system within military applications. As an example, the regulation framework presented in Figure 4-1 would be applied currently to certify a machine learning-based missile collision avoidance system in a cargo-transport aircraft such as the A400M.

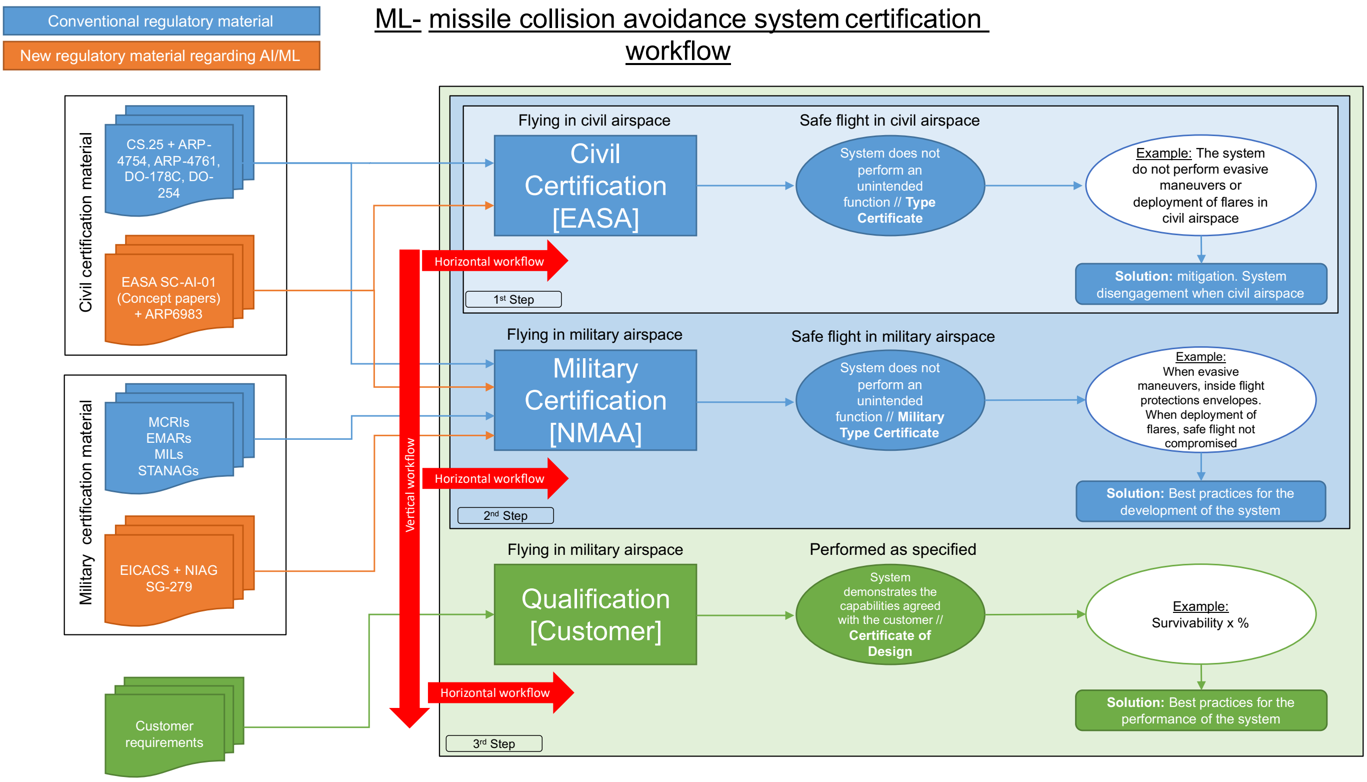

Figure 4-1: ML-missile collision avoidance system certification workflow (Source: Own elaboration).

Vertical Workflow:

- **Step One:** The aircraft undergoes certification for civil airspace by the EASA. This certification allows the aircraft to operate in civil airspace, including take-off and landing at civil airports, for potential civil cargo missions.

Example: In the case of implementing a missile avoidance system for the A400M, which operates in both civil and military airspace, it becomes essential to demonstrate to EASA that the system will not perform an unintended function. The system will not activate defence measures like flares or trigger evasive manoeuvres when flying in civil airspace. To mitigate this, one solution might involve disabling the system when the aircraft is within civil airspace.

- **Step Two:** Following civil certification, the aircraft then undergoes certification for military operations by the national military authority. The advantage here lies in reusing most of the type certificate obtained during the civil certification, with only a few modifications or additions necessary to tailor it for military operations.

Example: Regarding military certification, the primary objective is to ensure that the missile avoidance system functions as intended without compromising flight safety. This includes ensuring that evasive manoeuvres remain within the flight envelope protection and that defence measures like flares do not pose a threat to the aircraft itself. Since there is a trade-off between mission performance and safety, certain aspects of the civil type certificate may need adjustment. For example, the safety assessment for the missile avoidance system may be less stringent, recognising that it is more critical to be hit by a missile than to compromise safety slightly (within safe limits) and extend flight envelope protection.

- **Step Three:** The manufacturer must also adhere to customer-specific requirements, necessitating unique capabilities or performance standards.

Example: Qualification aspects may also be considered, with the customer requesting assessments of the system's survivability against various types of missiles and evasive manoeuvres, or the effectiveness of different flare types. These assessments could aim for specific survivability percentages, providing an additional layer of validation.

Horizontal framework flow

The three main horizontal workflows consist in identifying the meaningful certification material as input for each different process, identifying the authority and the flight space, and as an output, the certificate obtained within an example.

- **Civil horizontal workflow:** inputs civil materials as the one described in chapter 2.2.2 Airworthiness stakeholders and regulatory material and current AI/ML on going material or guidelines, identified in 3.2.1 AI/ML regulatory framework in civil domain. EASA as authority entity and type certificate (TC) as an output.

- **Military horizontal workflow.** Although reusing the civil type certificate, there are systems that haven't been certified in operation because in civil airspace are disengaged. As input, civil guidelines and some complementary military guidelines that civil material does not cover. In front of the national military authority or consortium depending on the platforms. As output, military type certificate (MTC).

- **Qualification workflow.** Customer requirements as an input, certificate of design (CoD) as an output. Does not concern airworthiness.

EASA concept paper objectives work as an initial guidance for AI-based systems in civil aviation. However, the military guidance material in terms of safety critical avionic systems is non-existing. Regarding the qualification, the objectives and requirements fall to the customer, and this is not in the scope of the thesis.

After having been placed in the framework and having seen an exact example of a system in that framework, the methodology with which this study is approached is as follows.

The methodology employed in this scientific work focuses, concretely, on the analysis of the feasibility of applying the objectives outlined in the EASA concept documents for civil AI/ML-based avionics to the field of military AI/ML-based avionics. This transferability assessment is conducted by examining how the EASA proposed guidelines can be applied to an AI-based military airborne system (see chapter Case study). Subsequently, a comprehensive gap analysis is conducted to identify gaps in guidance materials and specific objectives pertaining to AI/ML-based systems. It is intended to point out areas where EASA objectives may not be directly applicable or where adaptation for military purposes is essential.

In considering the airworthiness of military aircraft, a notable observation emerges regarding the availability of AI-related materials. While the objectives of the EASA concept paper provide an initial reference point for civil aviation, the military sector lacks comparable guidance materials.

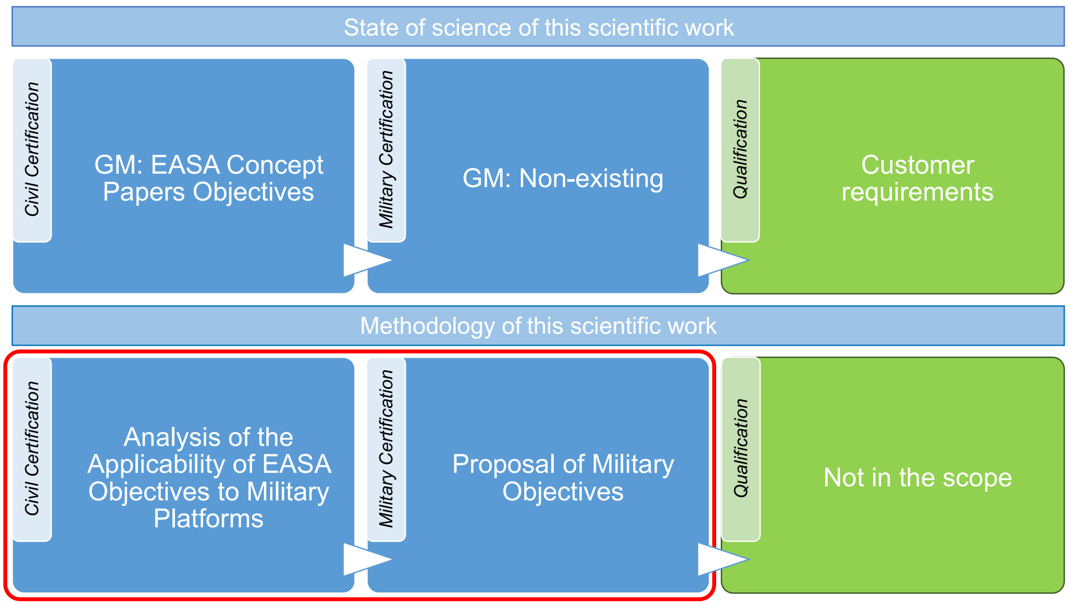

Figure 4-2: Methodology of this scientific work.

The methodology highlighted in red and presented in Figure 4-2, not only not only bridges the gap between civil and military certification but also paves the way for the responsible integration of AI into military operations. It acts as a catalyst for efficiency, knowledge dissemination, and the harmonisation of efforts, ultimately contributing to the advancement of military aviation in the age of AI.

5 Case study

As we review the multitude of potential applications within the scope of our research, it becomes clear that careful consideration is required to identify the ideal case study that will seamlessly align with our thesis objectives. To facilitate this selection, a set of meticulous criteria and requirements has been established. These criteria serve as a guiding framework to ensure that the selected case study not only meets the research objectives, but also reflects the broader landscape of applications outlined in the matrix. The following sections explain the importance of these criteria and their influence on the case study selection process:

I. Military application: The use case should have a direct or significant relevance to military operations or defence applications. This ensures that the case study, along with the rest of the project, is aligned with military objectives and requirements.
II. Supervised learning & offline learning: The scope of the EASA concept papers is limited to supervised learning techniques. This limitation will be removed in the next version of this guidance document through a planned extension to unsupervised and reinforcement learning. An analysis of a reinforcement case study such as missile collision avoidance would not be meaningful as the EASA objectives do not yet cover reinforcement learning.
III. Autonomy level 1 and 2: Same as the previous point. The scope of EASA's initial guidance encompasses autonomy levels 1 and 2, as per their classification.
IV. Safety critical: The use case should be safety critical, particularly in terms of airworthiness for military applications. A military system which is non-safety critical, has to do with performance demonstration rather than airworthiness.
V. Public information: The use case should involve publicly available information since this scientific work does not consider a non-disclosure agreement (NDA).
VI. Strategic alignment: The alignment ensures that the system or platform has the potential for dual certification, allowing it to be used in both military and civilian contexts. This alignment can lead to cost-sharing, broader market opportunities, and enhanced interoperability between military and civilian systems. Some examples are UAVs, transport military aircraft (A400M), MALE UAV, fuel tanker aircraft. Fighter jets have always been just military certified.

Some case studies are proposed and assessed whether they comply with the requirements or not:

Case study 1 – 'MALE UAV ISR'

Proposal of a medium altitude long endurance (MALE) UAV to perform ISR military operations. A MALE UAV ISR is a basic aerial component of modern militaries today offering real-time battlefield tracking and classifying ongoing actions and enemy positions. This case study focuses on the development of an onboard AI-based classification system capable of accurately identifying and classifying military-relevant objects, including trucks, military bases, and missile launchers, in real-time. The system to be analysed is the ISR system of the MALE UAV which contains the NN, not the UAV itself as an autonomous vehicle. The system includes a CNN which classifies weather it is a truck or a missile launcher for example. It is trained with labelled data (supervised learning) and since it just supports information analysis and does not suggest any decision-making, it is classified as 1A 'human augmentation'. In terms of criticality, it is a non-safety critical system since the system does not compromise the safety of the flight, but only has to do with performance demonstration.

Case study 2 – 'Visual air-to-air refuelling (AAR) guidance'

This case study will focus on a visual refuelling guidance system and will be studied in the following section 5.1 Visual Air-To-Air Refuelling Guidance System.

Case study 3 – 'Flight formation'

The flight formation case study refers to the coordinated and strategic arrangement of multiple UAVs during mission execution. This formation involves maintaining a specific geometric configuration or pattern among the UAVs to adapt to mission requirements and the dynamic conditions of the operational environment. The primary objective of UAVs flight formation is to increase mission effectiveness by leveraging the collective capabilities of multiple UAVs, and optimising various aspects such as extended coverage, surveillance angles, search efficiency, attack capabilities and operational resilience. The system to be analysed is the flight control system of the UAV which contains the ML algorithm. The system is trained through reinforcement learning, such as using multi-agent reinforcement learning algorithms [75]. As it is an autonomous flight, the system could be classified as 3 A/B "advanced automation" depending on the user's override power. In terms of criticality, it is a safety critical system since the system does compromise the safe flight of the vehicle.

Case study 4 – 'Missile collision avoidance system'

The hybrid missile avoidance system use case consists of two interconnected subsystems, each contributing to different stages of the missile avoidance process and both having the possibility to integrate ML items.

4.1. Missile Awareness subsystem: The first subsystem of the missile avoidance system is responsible for detecting incoming missiles using infrared cameras. This subsystem utilises supervised learning algorithms to process the infrared data and identify potential missile threats. Training data includes a vast dataset of missile signatures to enable the AI model to accurately differentiate between friendly and hostile objects in real-time. Regarding its level of autonomy, since it only classifies and does not suggest any action or decision-making, it is classified as 1A 'human augmentation'. It is considered a non-safety critical system since its unintended function behaviour does not compromise the safety of the flight but the safety of the mission.

4.2. Missile Avoidance subsystem: The second subsystem incorporates a more advanced approach with reinforcement learning techniques. This subsystem receives input from the missile awareness subsystem and other sensors, radars and so on, and assesses the likelihood of a missile colliding with the aircraft. If a collision risk is detected, the system proposes optimal manoeuvre strategies and automatically deploys defence aid systems (DASS). The proposed manoeuvres aim to optimise the aircraft's trajectory to minimise the probability of a collision. These algorithms learn from simulated scenarios and real flight data to make informed decisions about the best manoeuvres to avoid a missile.

- Countermeasures like evasive manoeuvres: The missile avoidance system communicates with the flight management system (FMS) to implement the recommended manoeuvres from the missile avoidance subsystem. If authorised by the pilot, the FMS automatically adjusts the flight trajectory based on the recommendations received. The subsystem establishes a link with the aircraft's FMS, which is a conventional software component responsible for controlling the aircraft's flight path. This subsystem is safety critical since proposed evasive manoeuvres out of the flight envelope protections (erroneous function behaviour) can cause the loss of the aircraft. This subsystem is classified as AI level 2B since there is collaboration between humans and the AI-based system.

- Electronic warfare: Interaction with DASS. Alongside manoeuvre suggestions, the missile avoidance system automatically deploys the defence aid systems like flares to further confuse or divert incoming missiles. In this case, the system is classified as level 3B AI, as the decision and action taken is autonomous and cannot be overridden by the end-user. This system is also classified as safety critical since the wrong deployment of flares can also cause hazardous damage in the aircraft.

As conclusion, when a missile threat is identified by the missile warner subsystem (1A) and the likelihood of collision is determined by the missile avoidance, it interacts with the FMS proposing evasive manoeuvres (2B) and/or gives the order to the DASS to deploy (3B) defence aids.

The following Table 5-1 presents a summary assessment of the proposed case studies:

Criteria	**Case study 1: MALE UAV ISR**	**Case Study 2: Visual Air-to-Air Refuelling Guidance**	**Case Study 3: Flight Formation**	**Case Study 4: Missile Collision Avoidance System**	
				4.1. Missile Warning Sub-System	**4.2. Missile Avoidance Sub-System**
Task	Detection and classification of trucks, enemy bases, and missile launchers	Visual guidance for air refuelling	Unmanned mission formation	Detection and classification of missiles	Calculation of the probability of impact, proposal of evasive manoeuvres, and deployment of defensive aids
AI technique	Computer Vision (CNNs)	Computer Vision (CNNs)	Multi-Agent RL	Infrared camera (CNNs)	Hierarchical RL
Learning	Supervised Learning	Supervised learning	Reinforcement Learning	Supervised Learning	Reinforcement Learning
Autonomy	1A	1B	3 A/B	1A	2B/3B
Criticality	Non-safety critical	Safety critical	Safety critical	Non-safety critical	Safety critical
Civil aviation previous cases studies	ACAS X	Visual landing guidance system (CoDANN)	no	no	no
Dual certification	yes	yes	no	yes	yes

Table 5-1: Case studies assessment.

As it can be observed from the table above, the case study that meets the specific requirements aimed for is **Case study 2 'Visual air-to-air refuelling guidance'**.

5.1 Visual Air-To-Air Refuelling Guidance System

Introduction

Aerial refuelling remains one of the most captivating manoeuvres in aviation. It involves a specialised transport aircraft like the Airbus A330 MRTT carrying fuel, which positions itself in designated airspace, an operation permitted in only a few sparsely populated areas in Germany due to its riskiness. These operations often involve refuelling multiple aircraft in a carefully planned sequence before the refuelling aircraft returns to its military base. In certain situations, it's possible to refuel two aircraft simultaneously, provided weather and airspace conditions allow for it.

The Airbus A330 MRTT, which stands for multi-role tanker transport, is a versatile aircraft. Besides its roles as a freighter or passenger aircraft, it can be converted into a flying intensive care unit and, as mentioned, function as a flying fuel tanker capable of carrying up to 111t of fuel. Currently, this military version of the A330-200 commercial model is in service with four air forces, including the Bundeswehr, and seven other countries and NATO have orders for it. It's worth noting that aerial refuelling is an exclusively military practice, as commercial aviation must adhere to stringent safety regulations, making it economically and practically unfeasible.

There are essentially two established methods for aerial refuelling. In the first method, a boom system is employed, where the refuelling aircraft positions itself in front of the receiving aircraft. The rear of the refueler is equipped with a relatively rigid boom that has limited manoeuvrability, allowing it to connect with the receiving aircraft's fuel tank. The process involves locking the boom and nozzles in place to complete an electrical circuit, initiating the pressurised fuel transfer. During refuelling, the receiving aircraft flies in close formation behind and below the tanker aircraft. Once the refuelling is complete, the valves are sealed, and the telescopic boom is retracted. The tanker boom operator or Air Refuelling Officer (ARO) engages the boom tip into a receptacle on the aircraft. The boom method relies on the ability of the ARO to manoeuvre the rigid boom within its limited range.

An alternative method involves using a probe and a drogue, resembling an oversized shuttlecock, attached to a flexible hose that trails behind the refueler. To refuel, the refueler positions itself in front of the receiving aircraft, and the receiving aircraft's pilot approaches from behind until it's just 20 metres away and guides the probe into the drogue. The airflow exerts force on the drogue, establishing a connection between the fuel probe tip and the valve, and the pressurised refuelling process commences. Throughout the process, an ARO oversees operations, adjusting the fuel delivery volume and, depending on weather and local conditions, the refuelling speed. Video and infrared camera systems on the fuselage assist in monitoring and ensuring a secure connection, even during night-time operations. The autopilot handles maintaining the short distance between the aircraft and making minor course adjustments. Ideally, aerial refuelling takes place in favourable weather conditions with clear skies, excellent visibility, and stable wind conditions.

Figure 5-1: Air-to-air refuelling boom vs probe-drogue method.

The choice between these methods depends on factors such as the aircraft's refuelling certification. The boom method relies on the refueler's ability to manoeuvre the rigid boom within its limited range, while the probe-and-drogue method requires the receiving aircraft's pilot to fly with greater precision. The boom method is generally faster due to its larger diameter, whereas the probe-and-drogue method allows for greater separation between the two aircraft, reducing the risk of collision, thanks to its 22-metre hose in front of the 11.6-metre boom. Depending on the refuelling method used, the transport aircraft can transfer up to 1,590 kg/min of fuel. This occurs at altitudes ranging from 1,500 to 10,000 metres, while both aircraft are flying at a speed of approximately 500 km/h.

Notably, the Airbus A330 MRTT has achieved the distinction of being the first tanker certified for automated aerial refuelling during daylight hours. To obtain this certification, Airbus developed the automatic air-to-air refuelling (A3R) system, which relies on automation to potentially reduce the workload of the ARO to merely monitoring the process. This system utilises automated image recognition and processing driven by information technology, enabling precise alignment of the boom tip with the receiving aircraft's intake, typically within a few centimetres. Real-time verification of correct alignment and the stability of the receiving aircraft is also possible. Importantly, the A3R system does not require any modifications to the receiver aircraft, enhancing the safety of refuelling operations. In the medium term, such automated refuelling processes are expected to operate with a high degree of autonomy, ensuring reliability even in challenging visibility and atmospheric conditions. The next phase, represented by the A4R project, aims to automate tasks carried out by the receiving aircraft. While this technology has the potential to alleviate some of the workload for the crew, planning and executing these operations will still necessitate substantial experience and aeronautical expertise. Nevertheless, it holds the promise of enabling aerial refuelling to take place in more challenging conditions than what pilots can currently manage.

The A330-MRTT is fitted with advanced avionics as well as fly-by-wire control systems featured in the civil A330 aircraft. It is also equipped with a customised suite of military avionics, like defence aids systems and electronic warfare. The mission systems are the Air-to-Air Refuelling (AAR) Systems. The AR systems are controlled from the Operator Console located in the flight deck. This arrangement facilitates pilot and ARO coordination. The Aerial Refuelling Boom System (ARBS) is operated by fly-by-wire controls and electrical actuators, together with a 3-dimensional (3D) video system. The 3D video system provides the operators with a clearer and realistic visualisation of ongoing AAR operations.

A typical tanker aircraft platform architecture is shown in Figure 5-2:

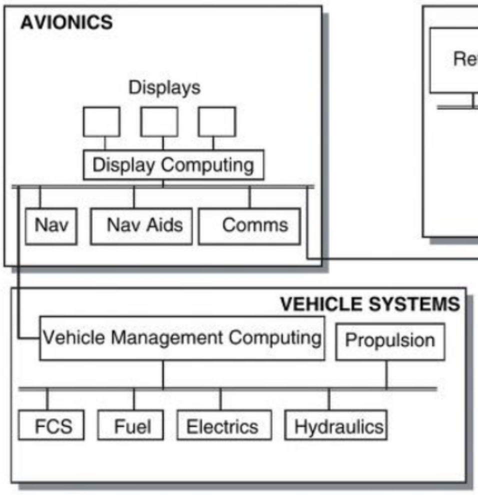

Figure 5-2: Typical air-to-air tanker platform architecture extracted from [22].

The objective of this case study is to develop a visual refuelling guidance system that uses ML to assist the ARO in aligning the boom with the fuel intake of the receiving aircraft. The system is expected to detect the receiving aircraft, classify the aircraft type, extract key points from the aircraft to determine its position and the position of the fuel intake, and finally provide boom alignment guidance, such as deviation correction, to the ARO.

Figure 5-3: Visual air-to-air refuelling guidance system [76].

This case study presents a potential application of ML in military use; however, it is an original creation, and does not involve any proprietary information or technology currently under development. The numerical values and parameters, while realistic, have been generated just for illustrative purposes. The case study system description will be based in the visual landing guidance system published by Daedalean and EASA in CoDANN I [5], and by Daedalean and FAA in [21] and the Runway Sign Classifier (RSC) published by the FSD institute in [57].

The system will be assessed, as previously described in 4 Methodology, following the "First usable guidance for Level 1 & 2 machine learning applications" [4] paper published by EASA as a certification guideline. Due to the limitations of the thesis, such as time and experience, this is not intended to be a complete certification exercise, but to provide an in-depth examination of some objectives. The proposed EASA objectives are divided into the following building blocks: trustworthiness analysis, AI assurance, human factor for AI, and AI safety mitigation. The latter are subdivided into sub-blocks as follows:

Building blocks	Sub-building blocks
Trustworthiness Analysis	Characterisation of AI
	Safety Assessment
	Information Security Assessment
	Ethics-based Assessment
AI Assurance	Learning Assurance
	Development Explainability
Human Factors	Operational Explainability
	Human-AI Teaming
	Modality of Interaction
AI Safety Risk Mitigation	AI Safety Risk Mitigation

Table 5-2: EASA AI building blocks.

The anticipated proportionality of this objectives is driven by the AI Level and the criticality (assurance level) of the item containing the ML model.

	Applicability by Assurance Level
●	The objective should be satisfied with independency.
○	The objective should be satisfied.
	The satisfaction of the objective is at the applicant's discretion.

Table 5-3: Assurance Level proportionality.

	Applicability by AI Level
	The objective should be satisfied for AI level 1A, 1B, 2A and 2B.
(green)	The objective should be satisfied for AI level 1B, 2A and 2B.
(orange)	~~The objective should be satisfied for AI level 2A and 2B.~~
(blue)	~~The objective should be satisfied for AI level 2B.~~

Table 5-4: AI Level proportionality.

In this chapter we are going to assess some of the building block objectives that affect our case study. Since we have classified our system as AI level 1B and following the applicability of the previous tables, the applicable objectives for only AI level 2A and 2B will not be considered. As for the assurance level, having classified the system as safety critical (we do not know yet which DAL) we can predict that most of the objectives should be met. DAL D or non-safety related systems present some objectives that could be satisfied at the discretion of the applicant. The objectives presented in the "First usable guidance for Level 1 & 2 machine learning applications" [4] paper and their classification regarding the two criteria presented in Table 5-3 & Table 5-4, can be found in Appendix B, Table 7-2: EASA objectives for AI level 1 & 2.

I. Trustworthiness Analysis

a. Characterisation of the AI application

Objective CO-01: The applicant should identify the list of end-users that are intended to interact with the AI-based system, together with their roles, their responsibilities, and their expected expertise (including assumptions made on the level of training, qualification, and skills).

Objective CO-02: For each end-user, the applicant should identify which high-level tasks are intended to be performed in interaction with the AI-based system.

Objective CO-03: The applicant should determine the AI-based system taking into account domain- specific definitions of 'system'.

Objective CO-04: The applicant should define and document the ConOps for the AI-based system, including the task allocation pattern between the end-user(s) and the AI-based system. Focus should be put on the definition of the OD and on the capture of specific operational limitations and assumptions.

<u>ConOps</u>

Before describing the system, to guarantee compliance with the objectives of the AI reliability guidelines, an overview of two ConOps (see Table 5-5) describing how the system will be operated is established, including the task allocation and operating conditions of the AI-based system.

	ConOps1	**ConOps2**
High-level task description	ARO advisory. The system assists the ARO in aligning the boom with the fuel intake by proposing real-time corrections for alignment deviations (left/right/up/down)	Full autonomy. The ARO monitors the AAR autonomously guided visual refuelling system which performs the deviation corrections autonomously
Purpose	Enhance ARO's visual perception	Completely Automated AAR Process.
End-user & and end-user interaction	ARO. Full control of the boom.	ARO. System oversight. **Elimination of ARO position as a future objective*
System interface	Mission Operator Display	Cockpit Display
Level of autonomy	1B 'Human assistance'	3A 'Overridable autonomy'
System criticality	Safety critical	Safety critical
Operating limits	Altitude: 1,500-10,000 m Speed: 500 km/h Retracted boom length: 11.6m Extended boom length: 18.20m Boom DoFs: two angular degrees-of- freedom (pitch and yaw), and one translational degree-of-freedom	
Key Technologies	Computer vision & augmented reality display	Computer vision, AI control algorithms & advanced collision avoidance

Table 5-5: ConOps.

However, the decision to opt for ConOps1, comes down to the fact that the case presents a level of autonomy within the scope of the EASA concept paper [4], which is the one which will be used to advise the case.

System description (ConOps 1)

The visual AAR guidance system facilitates the task of aligning the refueler's boom with the receiver's fuel intake and is composed of two subsystems:

1. Perception subsystem:
    - Sensor, such a high-resolution camera.
    - Pre-processing unit (traditional software): It reduces the resolution and normalises the image.
    - Image analysis (CNN): Detection of the likelihood of aircraft presence and normalisation of the aircraft key points.
    - Post-processing unit (traditional software): Employed for tracking and filtering the CNN output.
2. Actuation subsystem: The second subsystem operates based on the information provided by the perception subsystem, which includes determining the presence of an aircraft and, if the probability is sufficiently high, obtaining the key points coordinates of the aircraft and the fuel intake. This phase is responsible for executing the visual AAR guidance detailed (right, left, up, down corrections). It can operate autonomously or provide guidance to the pilot, depending on the ConOps. Since this scientific work primarily addresses certification concerns related to the application of ML, it will not be delved into further details about the actuation subsystem. It is assumed to be developed using conventional technologies.

Subsystem 1 is an AI-based subsystem while subsystems 2 is a traditional subsystem.

System architecture

The visual AAR guidance system architecture follows the decomposition of an AI-based system definition introduced by EASA definition [4].

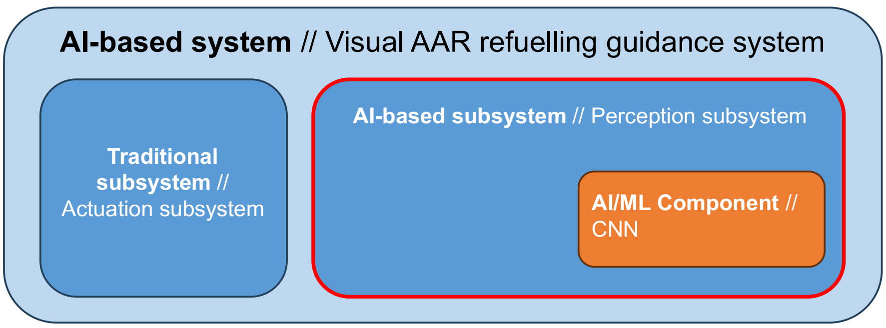

Figure 5-4: AI-based system definition by EASA [4] applied to this case.

System architecture

The process begins with the capture of images of aircraft using a camera, the captured images undergo pre-processing, which includes two main steps: resolution reduction where images are downscaled to reduce data size and accelerate processing, and pixel normalisation. This typically involves scaling pixel values to a specific range (e.g., between 0 and 1) or standardising them to have a mean of 0 and a standard deviation of 1. Next, a CNN is used to detect and classify the aircraft in the pre-processed images. The CNN is pre-trained using a dataset containing labelled aircraft images with their corresponding classes. Then a post-processing unit filters the output of the CNN detector and tracks the receiver. Then, another CNN extracts the normalised coordinates of key points on the detected receiver. These key points can be specific features of the aircraft, such as wingtips or the tail. The normalised coordinates of the key points are then post processed and sent to a pose converter. This component transforms the normalised coordinates of the receiver aircraft key points extracted from the image into the relative position of the aircraft and the fuel intake with respect to the boom tanker. Then, it computes the recommended corrections for the deviations found in the theorical alignment slope through 3 degrees of freedom (slope angle, deviation angle, and length of the boom).

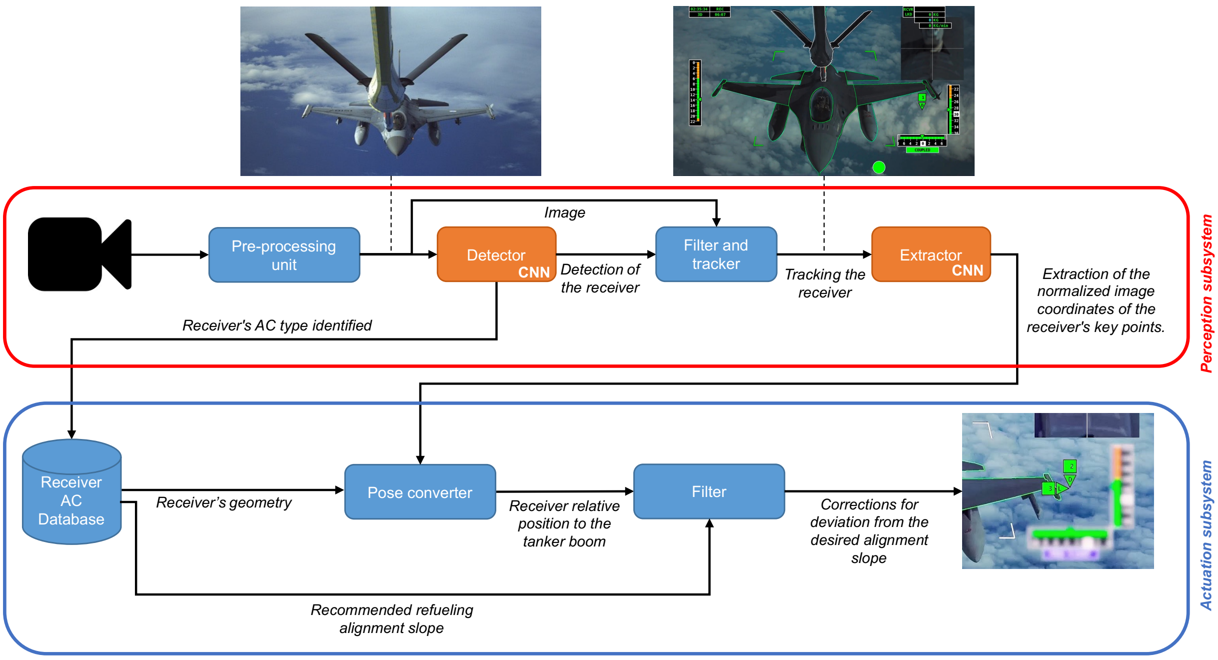

Figure 5-5: Visual AAR guidance system architecture (own elaboration).

In the following sections, the focus will be in the AI-based perception subsystem, highlighted in red in the previous Figure 5-4 and Figure 5-5.

Objective CO-05: The applicant should document how end-users' inputs are collected and accounted for in the development of the AI-based system.

The Kappa vision system conists of distinct subsystems designed to perform specific functions. The primary system offers a 3D perspective of the tank boom contact area through two separate optical channels, both delivering full HD quality. This visual input serves as the primary tool for the operator, while the central video management unit overlays pertinent data related to tank operations. A secondary system contributes to the overall situational awareness by combining imagery from three cameras, each offering a wide viewing angle of 240-330 degrees. This comprehensive panoramic view ensures continuous monitoring of operations, with integrated day and night vision capabilities. The systems exhibit a degree of redundancy, which enhances operational safety by allowing continued functionality in the event of system failures [76].

Figure 5-6: Kappa vision system [76].

Objective CO-06: The applicant should perform a functional analysis of the system.

A possible functional decomposition of the system is the following:

F1: Detect the Refuelling Receiver's Aircraft

This function involves using computer vision and ML – based perception to detect and track the refuelling receiver's aircraft in real-time.

- F1.1: Capture real-time imagery data.

Capture high-resolution imagery data of the external environment of the receiver aircraft.

- F1.2: Pre-process the image.

Pre-process the captured images to enhance clarity and decrease resolution.

- F1.3: Detect and classify the refuelling receiver aircraft in an image.

Implement AI-based algorithms to detect if there is an aircraft in the captured image and which type of aircraft. To output a crop of an image with the receiver aircraft in a bounding box.

- F1.4: Track the receiver aircraft in an image.

Implement a classical object tracker software to track the aircraft. Post-process and filtering the image.

- F1.5: Extract the normalised coordinates of the key points of the receiving aircraft in the given image.

Implement AI-based algorithms to detect the position of the receiver's aircraft key points in the given images.

F2: Provide Real-time Alignment Guidance

This function assists the ARO in maintaining precise alignment during the process of introducing the refuelling boom into the receiver intake.

- F2.1: Pre-process a cropped image assumed to contain the receiver aircraft.

Pre-process the data related to the receiver's aircraft normalised coordinates.

- F2.2: Compute the pose of the receiver's aircraft and its intake.

Use the aircraft type database to transform the normalised coordinates into the relative pose of the aircraft.

- F2.3: Provide alignment feedback.

Calculates the recommended alignment slope (slope angle, deviation angle, length of the boom). Then, it computes and outputs the lateral/vertical glidepath deviations from the intake alignment. Display real-time alignment feedback on the displays of ARO.

- F2.4: Adapt to changing conditions.

Continuously monitor and adapt alignment cues based on factors like turbulence or aircraft manoeuvres.

- F2.5: Ensure safety overrides.

Implement safety overrides to allow the disablement of the guidance.

F3: Monitor and Maintain System Health

This function focuses on monitoring the health and status of the visual guidance refuelling system.

- F3.1: Sensor monitoring.

Continuously monitor the sensors used for image capture and perception.

- F3.2: Image characteristics monitoring.

Monitor image quality and characteristics to ensure reliable data.

- F3.3: Internal system health.

Continuously monitor the internal health of the visual guidance system.

- F3.4: System component testing.

Test system components during power-up to identify any issues.

- F3.5: Enable/Disable output.

Determine whether the system's output should be enabled based on sensor and system health.

- F3.6: Abortion signal.

Signal when a refuelling manoeuvre should be aborted due to safety concerns.

The functional allocation to the subsystems and components can be done as follows:

Subsystems	Components	Functions
1 (perception)	Optical sensor	F1.1, 3.1
1 (perception)	Pre-processing unit	F1.2
1 (perception)	CNN Detector	F1.3
1 (perception)	Tracker/filter (post-processing unit)	F1.4
1 (perception)	CNN Extractor	F1.5
2 (actuation)	Pose converter	F2.1, F2.2
2 (actuation)	Filter	F2.3
Avionics display	Avionic display	F2.3, F3.6, F1.4
Main processing unit	Main processing unit	All the others

Table 5-6: Function allocation.

Objective CL-01: The applicant should classify the AI-based system, based on the levels presented in Table 2-2.

The AI Level 1B 'Human cognitive assistance' classification is justified by providing additional/advisory information (support to information analysis) to the ARO with suggestions for action or decision-making, such as right, left, up or down corrections.

b. Safety assessment of ML applications or compliance analysis

Objective SA-01: The applicant should perform a safety (support) assessment for all AI-based (sub)systems, identifying and addressing specifies introduced by AI/ML usage.

ID	Function Description	Failure Condition title	Phase flight	Effect of the failure on the aircraft, crew, and occupant	FC classification	Reason for FC classification
FC1-1	F1	**Total loss** of receiver’s AC detection function, **indicated** to the crew.	In flight	Loss of capability to detect the receiver’s AC. The ARO performs a visual alignment.	Minor	Classified as Minor due to slight increase in ARO's workload. He is trained for manual visual approach.
FC1-2	F1	**Total loss** of receiver’s AC detection function, **not indicated** to the ARO.	In flight	System does not alert to the loss of function, but ARO will realise this when trying to use the system. The ARO performs a visual alignment.	Minor	Classified as Minor due to slight increase in ARO's workload. He is trained for manual visual approach.
FC1-3	F1	**Erroneous false positive** receiver’s AC detection function: system is detecting something that is not an aircraft	In flight	The system detects a false positive and outputs a thumbnail of this false positive to the ARO. The ARO can confirm that the thumbnail does not match the actual receiver’s AC. The ARO does not allow the system to perform F2 (visual AAR guidance). ARO performs a visual alignment.	Minor	Classified as Minor due to slight increase in ARO's workload. He is trained for manual visual approach.
FC1-4	F1	**Erroneous false negative** receiver’s AC detection function: system does not detect an aircraft it should.	In flight	The system does not detect the receiver’s AC in view and in accordance with ConOps. The ARO communicates the receiver’s AC to attempt the approach again or ultimately performs a visual alignment.	Minor	Classified as Minor due to slight increase in ARO's workload. He is trained for manual visual approach.
FC2-1	F2	**Total loss** of visual ARR guidance function, **indicated** to the ARO	In flight	Loss of capability to provide alignment guidance. The ARO performs a visual alignment.	Minor	Classified as Minor due to slight increase in ARO's workload. He is trained for manual visual approach.
FC2-2	F2	**Total loss** of visual ARR guidance function, **not indicated** to the ARO.	In flight	System does not alert to the loss of function, but ARO will easily detect the loss of guidance. The ARO performs a visual alignment.	Minor	Classified as Minor due to slight increase in ARO's workload. He is trained for manual visual approach.
FC2-3	F2	**Erroneous** ARR guidance function,	In flight	The system provides erroneous alignment	Minor	Classified as Minor due to slight

		but **indicated** to the crew		guidance which is detected and alerted to the ARO. The ARO performs a visual alignment.		increase in ARO's workload. He is trained for manual visual approach.
FC2-4	F2	**Erroneous** ARR guidance function, but **not indicated** to the crew	In flight	The system provides erroneous alignment guidance that seems correct to the ARO. ARO will detect the erratic behaviour by visual confirmation, correct the approach path, and revert.	Major	Classified as Major due to significant reduction of safety margins. In the case of non-rapid reaction of the ARO severe effects if impact.
FC3-1	F3	**Total loss** of monitoring function	In flight	System monitoring is not working so there is no indication of performance or internal failures of the system. ARO is not aware if there is a problem with the system, and ultimately will only realise if there is an issue through visual confirmation or crosscheck with other aircraft systems.	Minor	Classified as Minor due to slight increase in ARO's workload. Disengagement of the system.
FC3-2	F3	**Erroneous** behaviour of the monitoring function	In flight	System monitoring starts outputting spurious failure messages and alerts. Flight crew will notice that the system is misbehaving and will need to take more effort cross checking the behaviour of the system or alternatively revert to an alternate system.	Minor	Classified as Minor due to slight increase in ARO's workload. He is trained for manual visual approach.

Table 5-7: Failure condition assessment.

Function	FDAL	Related FCs limitation
F1: Perception subsystem	D	FC1-1, FC1-2, FC1-3, FC1-4
F2: Actuation subsystem	C	FC2-4
F3: Monitoring subsystem	D	FC3-1, FC3-2

Table 5-8: FDALs allocation.

Since the functions are all from the same system, the higher FDAL defines the DAL allocation for our overall embedded visual AAR guidance system: **DAL C.** Nevertheless, the DAL allocation for our perception subsystem containing the ML component is **DAL D.**

The guidelines for conducting a safety assessment followed are adopted from the civil world, ARP4754A and ARP4761. Nevertheless, MIL-882E military standards for a military case stud, may increase the frequency of occurrence for the same severity level:

- ARP4754A and ARP4761
- **Military: MIL-882E**

Objective ICSA-01: The applicant should identify which data needs to be recorded for the purpose of supporting the continuous safety assessment.

For example: Inputs can be used to ensure that input distribution assumptions are verified, and outputs can be compared with GPS to identify discrepancies not explained by the level of accuracy of these systems. In this case, the GPS position of the receiving aircraft relative to the tanker can be compared to that extracted by the computer vision system.

c. Information Security Assessment

Even in the absence of inherent weaknesses in AI systems, the primary concern for the military today lies in the realm of adversary attacks. It is prudent to assume that potential adversaries will actively attempt to deceive or compromise any accessible AI system used in military operations. These attempts cover a wide range of tactics, such as tampering with image recognition engines and sensors, cyberattacks aimed at circumventing intrusion detection systems, and inserting altered data into logistics systems to disrupt supply chains with false claims. Adversarial attacks can be classified into four distinct types: evasion, inference, poisoning, and extraction. These forms of attack have been shown to be relatively simple to execute and may not even require advanced computer skills.

Evasion attacks focus on fooling an AI system, usually with the goal of avoiding detection. For example, this could be to hide a cyberattack or to convince a sensor that a military tank is a harmless school bus. In this context, the ability to remain hidden from AI sensors may emerge as a critical survival skill in future military scenarios. Consequently, innovative AI camouflage techniques may need to be developed to outmanoeuvre AI systems, as it has been shown that even simple methods such as strategic tape placement can effectively fool AI. Evasion attacks are often preceded by inference attacks, which aim to gather information about the AI system, subsequently exploited to facilitate evasion attacks.

Poisoning attacks, on the other hand, aim to corrupt AI systems during their training phase to achieve malicious goals. These threats can involve adversaries gaining unauthorised access to the data sets used to train our AI tools. For example, they could introduce mislabelled vehicle images to fool targeting systems or manipulate maintenance data to mask impending system failures as normal operation. Given the vulnerabilities in our supply chains, these scenarios are conceivable and would be difficult to detect.

Exfiltration attacks leverage access to an AI's interface to gain information about its operation, enabling the creation of parallel models. If our AI systems are not adequately protected against unauthorised users, these users could predict the decisions made by our systems and exploit these predictions to their own advantage. One can imagine a scenario in which an adversary predicts how an unmanned AI-controlled system will respond to specific visual and electromagnetic stimuli, thus influencing its route and behaviour [77].

Provision ORG-02: In preparation of the ~~Commission Delegated Regulation (EU) 2022/1645 and Commission Implementing Regulation (EU) 2023/203~~ applicability, the organisation should assess the information security risks related to the design, production, and operation phases of an AI/ML application.

The Commission Delegated Regulation (EU) 2022/1645 and Commission Implementing Regulation (EU) 2023/203 applicability do not apply to military operations.

In case of the visual AAR refuelling system case study, the adversarial attacks may not be that relevant since this operation does not include interaction with the adversary. Nevertheless, any cyber threat or vulnerability in the AAR visual guidance system could compromise the safe operation of the receiver, potentially leading to an impact and rendering it inoperable.

d. Ethic-based Assessment

Objective ET-01: The applicant should perform an ethics-based trustworthiness assessment for any AI-based system developed using ML techniques or incorporating ML models.

EASA has developed a comprehensive list of ALTAI questions adapted to aviation and AIRBUS has performed an initial assessment of the FCAS program with these guidelines (see FCAS chapter). However, the NATO principle of use should be considered, as the military domain is more constrained by ethics or flexible PRUs do not follow the EU ALTAI ethical guidelines because it is military.

1. **Lawfulness:** AI applications shall adhere to both national and international legal frameworks, including pertinent human rights and humanitarian laws.

European Parliament published on January 2021 a resolution report on artificial intelligence and international law titled “Questions of interpretation and application of international law in so far as the EU is affected in the areas of civil and military uses and of state authority outside the scope of criminal justice - P9_TA (2021)0009” [78]. It emphasises the need for the development of a common European framework to regulate these technologies, ensuring that they are human-centred and adhere to international humanitarian and human rights law. The document advocates for meaningful human control over AI and robotics, especially in the context of military applications. Furthermore, it calls for the development of a robust international legal framework, including an AI arms control regime, to govern AI's role in warfare and defence. The text highlights the importance of AI in maintaining peace, monitoring international security, and ensuring ethical research and design. It seeks to address the risks and implications of these technologies, urging vigilance in upholding human control, ethical principles, and international law in the face of technological advancements.

International law governs relations between states. The **Law of Armed Conflict (LOAC),** also known as the Law of War (LOW) or **International Humanitarian Law (IHL)**, is the part of international law that governs the conduct of armed conflict. LOAC is a set of rules which seeks, for humanitarian reasons, to limit the negative implications of armed conflict. It defines the rights and obligations of the parties to a conflict in the conduct of hostilities [25].

The United Nation (UN) has been discussing since 2014 [79] the topic **Lethal Autonomous Weapon Systems (LAWS).** A weapon is considered fully autonomous when it completes the decision cycle for target engagement on his own [79]This real-time execution cycle can be characterised by the sequence of operational activities Find / Fix / Track / Target / Engage / Assess, i.e., the so called F2T2EA-Cycle. Besides, the lawfulness of performing an attack is governed by the applicable **Rules of Engagement (RoE).** RoE are directives to military forces (including individuals) that define the circumstances, conditions, degree, and way force, or actions which might be construed as provocative, may be applied. RoE and their application never permit the use of force which violates applicable international law. Promoting responsible use of autonomous weapons systems while upholding humanitarian values is also an upcoming AI challenge in defence domains.

2. **Responsibility and accountability:** The development and utilisation of the AI applications shall involve prudent judgement and care, with clear human oversight to ensure accountability.

The European Parliament resolution also emphasises the critical need for human oversight and Meaningful Human Control (MHC) in AI systems, particularly in military applications, to ensure accountability and ethical decision making. In our case study:

- Responsibility: ARO as the person responsible for control.
- Accountability: explanations of the chosen actions. The CNN should be able to justify why it recommends the proposed alignment.

3. **Explainability and traceability:** AI should be transparent, incorporating review methodologies, sources, and procedures.

Guarantee of learning of AI-based systems equivalent to civilian applications.

4. **Reliability:** AI applicability must be rigorously tested and assured for specific cases through certification processes.

Guarantee of learning of AI-based systems equivalent to civilian applications.

5. **Governability:** AI applications should be aligned with their intended functions, enabling human-machine interaction and the capability to identify and address unintended consequences.

When failure conditions occur, ARO disengages the system and performs a manual visual alignment. Governability is more critical in the military domain, that is why the introduction of a mission operator in different platforms. Really high workload and complicated capabilities for just the flight crew.

6. **Bias mitigation:** Proactive measures should be taken to minimise unintended bias in both AI applications and data sets.

Assessment of threats biased based on limited data or faulty assumptions. If threat assessments are biased, it could lead to inaccurate judgments about the level of danger or the appropriate response. Bias mitigation efforts aim to identify and rectify any biases in threat assessments to ensure they are as objective and accurate as possible.

ALTAI [61] does not cover all the ethical aspects of AI systems in military. A qualitative assessment of which PRUs are covered by ALTAI guidelines is presented as follows:

NATO's PRUs	Covered with ALTAI
Lawfulness	○
Responsibility and Accountability	◐
Explainability and Traceability	●
Reliability	●
Governability	◐
Bias Mitigation	◐

Table 5-9: Level of coverage by ALTAI of NATO PRUs.

II. AI Assurance

The following section focuses on the perception subsystem containing the two ML components, the detector and the extractor CNNs.

a. Learning assurance

Learning assurance is based on system design and development. This is the previously mentioned paradigm shift between design assurance and learning assurance. From a software development model approach to a data driven approach. It consists of the W-shaped process proposed and counts with the following steps:

- Learning assurance process selection
- Requirements management and architecture management
- Data management
- Learning process management
- Training and learning process validation
- Learning process verification
- Model implementation
- Inference model verification and integration
- Data and learning verification
- verification of the subsystem requirements

Learning assurance process

Objective DA-01: The applicant should describe the proposed learning assurance process, taking into account each of the steps described in Sections C.3.1.2 to C.3.1.12, as well as the interface and compatibility with development assurance processes.

To achieve learning assurance, the W-shaped process proposed in the civilian world is followed (see Figure 2-5), however, some of the steps in this learning development/assurance process need to be examined more closely.

Requirements management

Objective DA-02: Documents should be prepared to encompass the capture of the following minimum requirements.

To continue, two tables with the main requirement of the perception subsystem 2 and the ML-components (CNNs) of this subsystem, respectively, are presented.

Perception subsystem requirements

ID	Description
AAR-1	The visual AAR guidance system shall detect and classify type receivers' aircraft according to training aircraft fleet dataset.
AAR-2	The visual AAR guidance system shall extract receivers' aircraft key points coordinates from the image.
AAR-3	The visual AAR guidance system shall operate under normal lighting conditions and with clear visibility.
AAR-4	The visual AAR guidance system shall recognise the receiver's aircraft to the up and down of the refueler located within a maximum angle of 20 degrees. (This means that the receiver must perform the approach and obtain the correct configuration to be in the correct range to activate the system*).
AAR-5	The visual AAR guidance system shall recognise the receiver's aircraft to the right and left and of the refueler located within a maximum angle of 35 degrees. (This means that the receiver must perform the approach and obtain the correct configuration to be in the correct range to activate the system*).
AAR-6	The AAR visual guidance system will recognise the receiver's aircraft if it is within 20 m of the refueler.
AAR-7	The visual AAR guidance system shall limit the detection latency to 500ms.
AAR-8	The visual AAR guidance system shall use a forward-facing camera with a frame rate of at least 10 Hz (10 fps).
AAR-9	The visual AAR guidance system shall have a precision of receiver detection of at least 95%.

Table 5-10: Perception subsystem requirements.

*If the receiver is not in a certain position (angle range up, down, right, left), the system will not detect the aircraft and guidance will not be provided. This is not because the detector is limited. We purposely limit the detector because the boom has limited manoeuvrability and if the aircraft is not positioned correctly, the contact will be erroneous and could impact the aircraft and fail to make successful contact.

Detector and extractor CNNs requirements

ID	Description
Det-CNN-1	The visual AAR guidance system detector CNN shall process the input 128 × 128 RGB image to detect the presence of the receivers' aircraft.
Det-CNN-2	The visual AAR guidance system detector CNN shall return the aircraft type of the detected receiver.
Det-CNN-3	The visual AAR guidance system detector CNN shall have an aircraft detection and aircraft type identification accuracy of at least 95%*.
Det-CNN-4	The visual AAR guidance system detector CNNs shall return the image with the receiving aircraft in it and its bounding box position.
Ext-CNN-5	The visual AAR guidance system extractor CNN shall extract the normalised coordinates of the receiver's aircraft key points in the RGB image.
Det&Ext-CNN-6	The visual AAR guidance system CNNs shall process the image within 100ms (<500ms).

Table 5-11: Detector and extractor CNNs items requirements.

*The visual AAR guidance system CNN identifies a different type of aircraft; the visual AAR guidance will be erroneous.

> **Objective DA-03:** The applicant should describe the system and subsystem architecture, to serve as reference for related safety (support) assessment and learning assurance objectives.

The (sub-)system's high-level architecture is described in Figure 5-5. The CNNs architecture description is not shown because it has not been implemented (number of layers, etc..).

Data management

> **Objective DM-01:** The applicant should define the set of parameters pertaining to the AI/ML constituent ODD.

Considering the following hypothesis:

- The receiver's aircraft and the tanker are stationary: same constant horizontal and vertical velocity. The refuelling operation starts when this configuration is achieved.
- Both aircrafts are in cruise configuration (their frames of reference are parallel) and aligned with the vertical and horizontal plane.
- The camera does not move with the boom, it is fixed to the aircraft. Consider a camera aligned with the horizontal plane and vertical plane of the aircraft.

- The boom does not roll. The refuelling boom is modelled as a rigid, telescoping rod with two angular degrees-of- freedom (pitch and yaw), and one translational degree-of-freedom. The boom is attached to the tanker aircraft with a resolute joint.

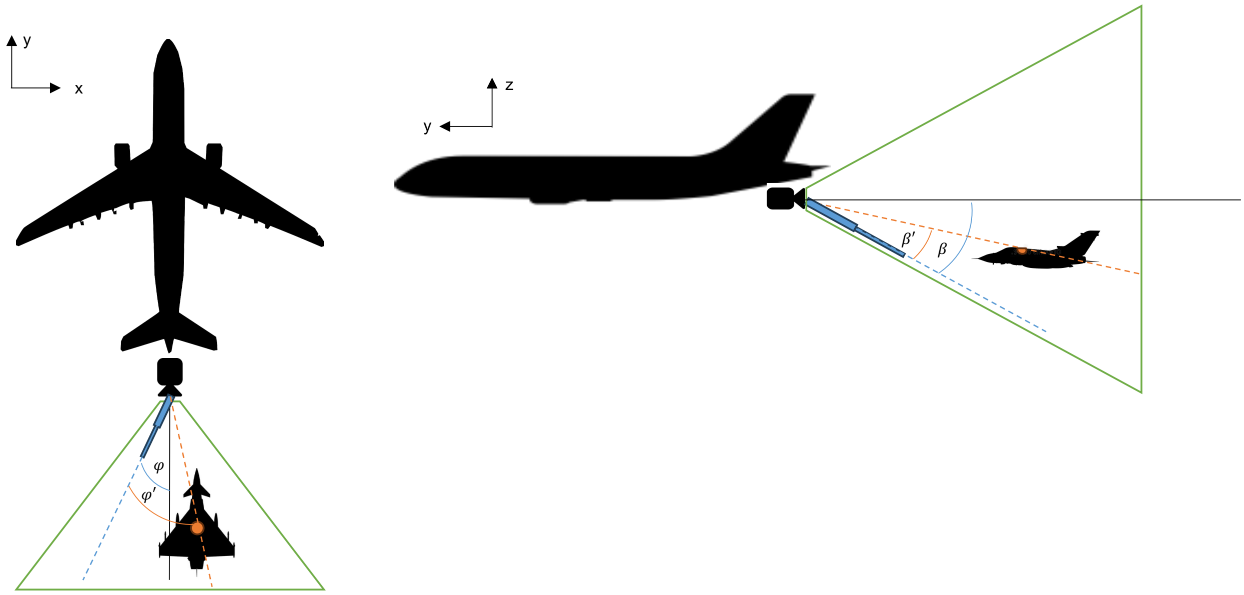

Figure 5-7: Relevant angles (Source: Own elaboration).

The relevant operating parameters include the following:

- **Alignment slope:** the orange dash line in Figure 5-7.
- **Boom pitch, $\boldsymbol{\beta}$:** A vertical angle between the camera direction view and the horizontal plane of the boom.
- **Boom yaw, $\boldsymbol{\varphi}$:** A horizontal angle between the camera direction of view and the vertical plane boom.
- **Effective slope angle, $\boldsymbol{\beta'}$:** The vertical angle between the alignment slope and the horizontal plane of the boom.
- **Lateral deviation angle, $\boldsymbol{\varphi'}$:** The horizontal angle between the alignment slope and the vertical plane of the boom.
- **Distance to the receiver's intake, $\boldsymbol{d}$:** Distance from the edge of the boom to the intake along the slope. Third degree of freedom.
- Time of the day: sun elevation angle and azimuth
- Weather

Objective DM-06: The applicant should identify data sources and collect data in accordance with the defined ODD, while ensuring satisfaction of the defined DQRs, in order to drive the selection of the training, validation and tests data sets.

The data sources are an integral part of the AI/ML project and can originate either from within the project applicant's organisation (internal sources) or from external entities. External sources can encompass open-source data or data acquired through contractual agreements established between the applicant and the data provider. In this case study, data sourcing and data collection present no significant issues because the data used to train the CNN comes from the internal organisation's own fleet (all kind of receiver's aircraft types). The data remains within the control of the interested party, and there is no compromise of confidentiality, as it's an internal resource. The labelled data used for model training pertains to various receiver aircraft to be refuelled during operations.

However, it's essential to recognise that in other defence use cases, where some data may be highly confidential, certain requirements are non-negotiable. Defence applications often involve numerous stakeholders, including nations, manufacturers, and governments. Addressing these concerns typically requires solutions such as federated learning, which ensures that data remains secure and does not leave the control of the contributing parties.

Federate Learning

Federated learning, referred also as collaborative learning, represents a decentralised approach to machine learning model training. This method avoids the need to transmit raw data from client devices to global servers. Instead, it relies on utilising raw data locally on edge devices for training, thus significantly enhancing data privacy. The final model is collaboratively constructed by aggregating these local updates.

Federate learning stands for three main pillars:

- Privacy: In contrast to conventional methods where data is sent to a central server for training, federated learning allows local training on edge devices, effectively preventing potential data breaches.
- Data Security: Only encrypted model updates are shared with the central server, ensuring robust data security. Secure aggregation techniques, like the Secure Aggregation Principle, enable the decryption of only aggregated results, bolstering data protection.
- Access to Heterogeneous Data: Federated learning ensures access to data distributed across various devices, locations, and organisations. This approach enables the training of models on sensitive data, such as defence data, while upholding the highest standards of security and privacy. Additionally, the diverse data sources enhance the generalisability of the models.

An initial standardised model is stored on the central server. Copies of this model are distributed to client devices, which then undergo training based on the data they generate locally. Over time, the models on individual devices become personalised, leading to an improved user experience. In the subsequent phase, the updates, which consist of model parameters, are securely shared with the central server from the locally trained models. The central model amalgamates and averages these diverse inputs to generate new insights. Given the diversity of data sources, this approach greatly enhances the model's potential for generalisation. After retraining the central model with these new parameters, it is once again shared with the client devices for the next iteration. With each cycle, the models accumulate a diverse set of information, leading to continuous improvement without compromising data privacy.

However, this process is not without its critics, mainly regarding the effectiveness and generalisability of the universal federated model. Typically, individual machine learning algorithms show sensitivity in handling the available data and discovering the intricate relationships that characterise them. Consequently, the predictions produced by such models may manifest substantial biases.

Another key concern in defence applications is not only confidentiality, but also data availability. Sometimes, it is necessary to train models against potential adversary systems, such as anti-missile awareness. To address issues related to the lack of completeness or representativeness of data, existing data can be supplemented with synthetic data, which can be derived from models, digital twins, or virtual sensors.

As mentioned above, this leads to another issue that must be considered, data poisoning. In scenarios where adversarial actors attempt to purposefully manipulate their own data to serve as training data for the future to undermine model performance.

Objective DM-12: The applicant should distribute the data into three separate and independent data sets which will meet the specified DQRs:

- the training data set and validation data set, used during the model training.
- the test data set used during the learning process verification, and the inference model verification.

The data is divided into training, validation, and test data set. Prevent that the data with which the model has been trained, is included in the validation or test data sets (see Figure 2-4 for each data set role).

Learning management

Objective LM-02: The applicant should capture the requirements pertaining to the learning management and training processes, including but not limited to:

- model family and model selection;
- learning algorithm(s) selection;
- cost/loss function selection describing the link to the performance metrics;
- model bias and variance metrics and acceptable levels;
- model robustness and stability metrics and acceptable levels;
- training environment (hardware and software) identification;
- model parameters initialisation strategy;
- hyper-parameters and parameters identification and setting;
- expected performance with training, validation and test data sets.

This high-level case study pertains to a proposed military application, and as of now, the development, training and implementation of the model has not been undertaken. Consequently, specific aspects such as model selection, learning algorithms, metrics, model environments, parameters, and hyperparameters have not been addressed. Nevertheless, it is anticipated that in terms of managing the learning process, the approach will be comparable, if not identical, to the civil approach.

Learning process verification

Objective LM-09: The applicant should perform an evaluation of the performance of the trained model based on the test data set and document the result of the model verification.

Model implementation

Objective IMP-03: The applicant should plan and execute appropriate development assurance processes to develop the inference model into software and/or hardware items.

Inference model verification and integration

Objective IMP-06: The applicant should perform an evaluation of the performance of the inference model based on the test data set and document the result of the model verification.

Independent data and learning verification

Objective DM-14: The applicant should perform a data and learning verification step to confirm that the appropriate data sets have been used for the training, validation and verification of the model and that the expected guarantees (generalisation, robustness) on the model have been reached.

For the previous steps of the W-shaped process, since the ML model is not implemented and the subsystem is assigned a level D, we propose to follow the ML development workflow elaborated by the FSD Institute at TUM in [56] for low criticality airborne applications (previously mention in 3.2.1.4 Industry (IPC) and academy). In this workflow, several assumptions are made to be able to apply standard activities and methods that comply with conventional standards, such as DO-178C. These assumptions are:

- To focus on ML components only. For traditional non-ML software development, DO-178C applies without modifications.
- To consider non-adaptive ML systems.

- To consider a deep neural network trained through supervised learning as a representative example.
- To implement ML model training at system level and to integrate pre-trained models into the software development cycle. Typical ML algorithms, such as DNN, are primarily composed of simple arithmetic operations, making them suitable as the foundation for low-level software requirements, which can be readily transformed into source code. In conclusion, ML models are considered low-level requirements. For level D software, traceability objectives of low-level requirements do not apply. It can be argued that testing ML models with test data sets helps with some verification objectives. This is like the simulation approach for model-based systems discussed in DO-331. This represents a mitigation of an incompatibility of ML-based systems and conventional standards. Nevertheless, NNs may require supplementary low-level requirements presented in a traditional format for tasks like specifying mathematical libraries for matrix operations. For example: to consider training dataset as ML model requirements.
- To apply the design assurance principles of the other applicable standards, such as ARP-4574A and DO-254, since compliance can be demonstrated using the same approach that for DO-178C.

With these assumptions our DAL D perception subsystem could complete the certification exercise following the conventional standard. This will cover the AI assurance issue, but not the explainability and HAII issues presented above.

b. Development/post-ops explainability

Objective EXP-01: The applicant should identify the list of stakeholders, other than end-users, that need explainability of the AI-based system at any stage of its life cycle, together with their roles, their responsibilities, and their expected expertise.

Objective EXP-02: For each of these stakeholders (or groups of stakeholders), the applicant should characterise the need for explainability to be provided, which is necessary to support the development and learning assurance processes.

This focuses on the list of stakeholders other than the end-users, as these have been identified already as per Objective CO-01, and their needs of explainability. Stakeholders usually need a deeper level of insight in the design details of the AI-based system,

This building block would follow the same process as in the commercial aircraft: identifying the stakeholders like the ones previously mentioned in Figure 2-8, their different needs of explainability and the methods to achieve this. Authorities, engineers, and customers do not have the same type of need; the system is transparent and accountable to the satisfaction of all involved parties.

III. Human factors for AI

a. Operational explainability

Objective EXP-05: For each output of the AI-based system relevant to task(s) (per **Objective CO- 02**), the applicant should characterise the need for explainability.

EASA anticipated operational explainability modulation		
AI Level	**Level 1B ‘Human assistance’**	**Visual AAR guidance**
Overall Assessment	The implementation of an AI-based system is expected to impact the current operation of the end-user with the introduction of, for example, a cognitive assistant.	Visual assistance
HAII Expected level of evolution in the human-AI interaction (HAII) compared to existing interactions	Medium change: There is a need for explainability so that the end-user can use the AI outcomes to take decisions/actions.	The system proposes actions and decisions such as corrections of the boom alignment as right, left, down, and up but the ARO decides whether to perform them or not.
Explainability Expected level of explainability needed during operation	Explainability is there to support and facilitate end-user decisions. At this level, the decision still requires human judgement or some agreement on the solution method.	An effective way to demonstrate explainability in a visual air-to-air refuelling guidance system is through the mission flight display. This display can showcase a magnified view of the receiving aircraft and its clearly defined key points (including the fuel intake), which are detected by the system. The display also shows the movement of the boom allowing the ARO to assess whether the suggested deviations from the alignment slope in all directions (up, down, right, and left) make sense. This visual representation serves as a powerful explanation of the output quality.
Guidance Need for specific human factors certification guidance linked with the introduction of AI-based systems	Specific guidance needed. Need for operationalising the explainability concept in the frame of future design and certification.	Definition of attributes of explainability with design principles: - Understandable explainability - Level of abstraction - Timeliness of explainability

Table 5-12: EASA anticipated operational explainability modulation.

Objective EXP-06: The applicant should present explanations to the end-user in a clear and unambiguous form.

The flight display shows a zoomed-in inlet of the receiver aircraft and its delimitation (unition of the key points), as detected by the system. The design of the system implies that if the key points are precisely positioned, then the guidance will be accurate. This provides the end-user with a powerful explanation of the quality of the output, in addition to the provided measures of uncertainty (see Figure 5-3). The display also shows the movement of the boom allowing the ARO to assess whether the suggested deviations from the alignment slope in all directions (up, down, right, and left) make sense. This visual representation serves as a powerful explanation of the output quality.

Objective EXP-07: The applicant should define relevant explainability so that the receiver of the information can use the explanation to assess the appropriateness of the decision / action as expected.

This helps the receiver understand the purpose and objectives behind the system's recommendations. For AAR, this could include goals related to maintaining safe distances, alignment accuracy, and refuelling efficiency. Explain the reasoning behind the AI-based system's decisions. This may involve outlining the logic, causal relationships, or the specific factors considered in making recommendations. In the context of AAR, this could involve detailing the factors affecting alignment, distance, and stability during the refuelling process. Provide detailed information on the contextual elements considered by the AI-based system when making decisions or implementing actions. For AAR, this may include data related to aircraft positions or weather conditions. Sharing this information allows the end-user to better understand and evaluate the system's recommendations. If the AI-based system is making trade-offs between operational needs, or risk analysis in the AAR process, ensure that these strategies are explained in the context of the decision. This helps the receiver understand the decision-making process and the reasons behind it. Clarify the sources of data used by the AI-based system for decision-making. In the case of AAR, this could involve specifying the data sources for aircraft positions, weather information, and other relevant parameters. It's crucial for users to know which data sources were used and their reliability, especially in a multi-crew environment.

b. Human AI teaming

As per table in Annex B the objectives of this sub-building block do not apply to autonomy level 1B.

c. Modality of interaction

As per table in Annex B the objectives of this sub-building block do not apply to autonomy level 1B.

In cases where there exists mission operators due to the exceptionally demanding workload of the flight crew and the complexity of the mission, the inclusion of a third user brings a need to reassess autonomy levels and the criticality of failure conditions. It is essential to also consider the modes of interaction among these three users. Additionally, we must account for the various end-users and their distinct requirements for levels of comprehensibility and explanation depth.

IV. Risk mitigation

SRM-01: Once activities associated with all other building blocks are defined, the applicant should determine whether the coverage of the objectives associated with the explainability and learning assurance building blocks is sufficient or if an additional dedicated layer of protection, called hereafter safety risk mitigation (SRM), would be necessary to mitigate the residual risks to an acceptable level.

SRM-02: The applicant should establish SRM means as identified in Objective SRM-01.

Some means of risk mitigation include monitoring of the output of the AI/ML constituent and passivation of the AI-based system with recovery through a traditional backup system (e.g., safety net) or when relevant, the possibility of the end-user to switch off the AI/ML-based function to avoid being distracted by erroneous outputs.

For instance, in our case study, as stated in the safety assessment, in ConOps 1, the end-user, ARO, has the capability to deactivate guidance and opt for manual visual guidance. In the scenario of future full autonomy, where there is no ARO, and the pilot's workload becomes too burdensome, it becomes imperative to oversee the system's output and have a robust traditional backup system in place.

Allied Tactical Publication (ATP) 3.3.4.2 “Air- to- Air Refuelling” published by NATO [80] states as a mitigation procedure:

“The air refuelling boom is controlled by a fly by wire control system. Should this system suffer certain failures; ARO may not be able to control the boom in one or more axis of movement. If certain failure occurs, coordinated action between the receiver pilot and the ARO will be required to prevent the boom from striking the receiver. The ARO may inform the receiver pilot by stating “Tanker Boom Fault” and initiate a disconnect. Immediately after that statement and a safe disconnect the receiver MUST move down and aft.”

6 Discussion

This chapter presents an analysis of the research findings and their implications in the context of the role of AI in military avionics systems. We will propose a solution to the gaps identified in the case study and discuss the research questions, how we have addressed them during the thesis and what results have been obtained.

6.1 Gaps Analysis

The following Table 6-1 addresses the gaps identified during the applicability of EASA concept papers for Level 1&2 ML applications to an AI-based military system and classifies them in the different EASA ML building blogs. It is essential to note that not all objectives were addressed, mainly due to the complexities involved and the time constraints inherent in conducting a complete certification exercise. The system in question has not been fully implemented but remains as a proposal. We therefore focused on addressing the identified objectives at a high level, with an emphasis on highlighting those aspects that would require additional layers or considerations to make them suitable for the military domain.

It is assumed that the elements that do not contain any gaps found, have been treated equally but do not differ from the civil guidelines.

EASA building blocks	Sub-building blocks	Case study traceability	Gap identified in military	STANAG section
Trustworthiness analysis	Characterisation of AI	I.a Characterisation of the AI application	Difficulty to define a unique ConOps regarding specific military aircraft roles. Regulatory framework generalisation issue.	I
	Safety Assessment	I.b - Safety assessment of ML applications or compliance analysis	Less restrictive. Trade-off between mission accomplishment and safety.	II
	Information Security Assessment	I.c - Information Security Assessment	Higher probability of having adversarial attacks from the enemy in different ways. More vulnerable.	III
	Ethics-based Assessment	I.d - Ethic-based Assessment	Compliment with extra humanitarian and conflict laws.	IV
AI Assurance	Learning Assurance	II.a - Learning assurance	Classified data, and/or not available real data from the adversary. Confidentiality and data privacy.	V
	Development Explainability	-	-	-
Human Factors for AI	Operational Explainability	III.a - Operational explainability	Mission displays / Heads-up display	VI
	Human-AI Teaming	III.b - Human AI teaming	Different AI-end-user interaction and workload.	VI
	Modality of Interaction	III.c Modality of interaction	Introduction of the mission operator. More complex interaction.	VI
AI Safety Risk Mitigation	AI Safety Risk Mitigation	-	-	

Table 6-1: Gaps analysis.

6.2 Proposal of a STANAG

In response to the unique challenge posed by the integration of AI in military systems, we propose the development of a STANAG with the aim of completing the civil guidelines. As the main outcome, civil guidelines are necessary but not sufficient. While civilian guidelines apply completely, they may not cover all specific military requirements. Our STANAG proposal is designed to comprehensively complete the ML objectives set by EASA ensuring that these guidelines are applicable in both civil and military contexts. This dual applicability is crucial to harmonise standards and achieve mutual recognition on both sides, streamlining the certification process. A significant advantage of this approach is that it facilitates dual certification, which can save time when certifying aircraft for military use.

To fill the previous gaps identified in Table 6-1, the following aspects should be covered by the proposed STANAG:

I. Regulatory framework generalisation

In the context of certification of AI/ML-based systems for military aircraft, depending on the military role of the aircraft and its operational context, certification criteria need to be adjusted and require specific considerations. Consequently, following the example of Figure 2-6, we propose to classify military avionics into three distinct categories:

- General Avionics: These systems support safe aircraft operation and include communication, navigation, flight control, cockpit displays, and safety systems. This category will follow the same approach as the commercial aircraft avionic systems.
- Mission Systems: These avionics are mission-specific and can include radar, electronic warfare, sensors, reconnaissance equipment, and data link systems.
- Weapon Systems: Avionics for weapons control and delivery, including weapons management, targeting, fire control, and countermeasures.

The adaptability of certification is crucial to ensure that the specific needs of each category of avionics are addressed. In a combat aircraft, consideration of ethics and compliance with LAWS and RoE are of paramount importance. In this environment, every decision made by the AI/ML system must be carefully evaluated to ensure that it conforms to the ethical and legal constraints governing the use of military force. Respect for international laws and rules of engagement is essential to avoid unintended consequences and maintain the legitimacy of military operations. Operational explainability, such as in the context of the interaction between the AI/ML system and the pilot via a Head Mounted Display (HMD), is another critical aspect of combat aircrafts. Given the high level of workload and rapid decision-making required in a dogfight, the pilot's ability to understand and trust the system's outputs is essential. The HMD can play a key role in displaying data and communicating critical information to the pilot. However, it is important to note that the level of explainability required may vary depending on the complexity of the task and the operational situation. In a dogfight, where speed and accuracy are crucial, the pilot may not have the time to delve into a detailed explanation of every decision made by the AI/ML system. Instead, it is essential that the system provides clear results and actions that the pilot can immediately understand and use to his advantage without the need for lengthy analysis.

On the other hand, in surveillance aircraft, where the primary focus is on data collection and threat identification, certification must focus on effective data collection and processing. AI/ML system training in the absence of enemy data is a major challenge that needs to be addressed. In situations where the interaction is primarily with a mission operator rather than a pilot, a different autonomy classification may need to be considered. This is because the impact of an AI/ML system on the pilot is different from its impact on the mission operator. This may also require the definition of specific safety and performance standards for this different interaction.

The proposed visual Air-to-Air Refuelling (AAR) guidance system, as outlined in the case study, falls within the category of mission systems avionics. Unlike weapon systems, this avionic system is not associated with offensive countermeasures, and therefore, ethical considerations related to the use of lethal force are not as pertinent in this context. Additionally, data availability is not a significant concern because the system's primary function is to facilitate the in-flight refuelling of military aircraft within the same fleet, which means that the data is readily accessible without privacy or availability issues. However, the key challenge in certifying this visual AAR guidance system lies in the delicate balance between mission effectiveness and flight safety. Given that the system's configuration is critical to the success of the refuelling operation, ensuring that it performs its mission without compromising safety is of paramount importance. In a military aircraft with this AAR system, there may be an AAR operator who is a mission specialist, and their role is to oversee and manage the refuelling process. This specialised role creates a new way of interaction between the AI/ML system and the flight crew. It reduces the workload on the pilot, who can focus on other critical aspects of the mission. This reduced pilot workload means that the operational explainability of the AAR system may not need to be as extensive as in civilian contexts, where the pilot is typically responsible for a broader range of tasks. In summary, the certification of the visual AAR guidance system, which falls within the mission systems category, comes with unique considerations.

In summary, certification of AI/ML-based systems on military aircraft is a complex challenge that requires an adaptive approach and consideration of the specific needs of each type of avionics. The proposal to classify avionics into different categories and the adaptation of certification criteria are important steps towards improving the safety and effectiveness of these systems in the military context.

II. Safety Assessment and Mission Accomplishment

The STANAG will provide guidance on conducting safety assessments that carefully balance reducing the likelihood of risk with the imperative of meeting mission objectives in the military context. It will address the trade-offs between safety and mission success, with a focus on ensuring that mission-critical decisions remain sound. If civil safety limits cannot be achieved, the MIL-STD 882E can be applied, refer to 2.2.3 System safety approach.

III. Poisoning and threat resilience

A central component of the STANAG will be strategies and measures to strengthen the resilience of AI systems against poisoning attacks and other adversarial threats. This will include guidelines for secure data handling, training, and validation to mitigate the risk of AI manipulation. It will follow a similar approach as the civil world, but it will consider more specific and remote scenarios characteristics of military missions.

IV. Ethical and legal considerations

Building on existing legal frameworks, the STANAG will incorporate specific ethical guidelines for military AI deployment. These guidelines will align with established NATO principles and address the unique ethical considerations that arise in military aviation.

V. Data collection, privacy, and security

STANAG will emphasise the importance of representative, complete and correct data collection for AI training. It will also consider the inclusion of adversarial scenarios in training data. In addition, it will address privacy issues, especially when dealing with classified information. It will propose new strategies to address this issue like federate learning.

VI. Human AI teaming and autonomy levels

STANAG will explore the introduction of mission operators based on the specific role of the military aircraft. It will also examine the possible introduction of new AI autonomy levels to suit different mission requirements, ensuring seamless collaboration between humans and AI. The introduction of mission displays, or HMD should be considered and respective specific standards regarding interoperability should be accounted.

7 Conclusions

7.1 Objectives fulfilment

The research has been guided by a set of well-defined research questions, each tailored to investigate a specific aspect of AI certification in military avionics. This chapter discusses each of these, how they have been addressed, and their conclusions.

RQ1. Knowledge compilation: Chapter 3.2 AI/ML regulatory framework outline a detailed analysis of ongoing efforts and initiatives to standardise AI technology in civil and military aviation, with a particular focus on avionics. In civil aviation, efforts are underway to standardise AI technology through the EU AI Act, collaborations with aviation regulatory bodies like EASA and the FAA, and the development of industry standards for ML applications. Interim certification solutions and research projects are also contributing to innovation. In military aviation, the adaptation of civil aviation guidelines and procedures for AI military airborne systems is occurring, but challenges arise due to national regulation and mission-specific requirements. Initiatives like the EICACS and the NIAG-SG279 are promoting AI standardisation in the defence sector. This analysis allows for an assessment of the current standardisation landscape and points out areas of alignment and divergence between the civilian and military sectors.

RQ2. Identification of AI applications in military avionics: Chapter 2.1.5 AI applications in military avionics introduces the multiple applications of AI in military avionics, such as enabling autonomous flight, performing autonomous ISR tasks, optimising mission execution, automating target detection and identification, providing real-time decision support to airmen, improving C4I operations, enhancing cybersecurity and resilience, facilitating training and simulation, predicting maintenance needs, and improving ATM. These applications not only improve operational efficiency and effectiveness, but also play a crucial role in ensuring the safety and success of military aviation missions. A breakdown of these main military domains in AI-based avionics systems, their benefits, which AI techniques can be mainly used, and their level of criticality can be seen in Appendix A: Classification of military AI-based airborne systems.

RQ3. Certification Challenges Assessment: Chapter 2.3 Challenges of certifying AI/ML-based systems in military aviation identifies the unique challenges associated with the certification of systems employing ML in military aviation, distinct from the common challenges encountered in civil aviation. Certifying AI/ML systems in military aviation poses multifaceted challenges, encompassing the need for robust security and protection against cyber threats, adapting to complex and dynamic operational environments, navigating strict export control regulations for sensitive technologies, ensuring integration into certified hardware, seamless interoperability with existing systems, maintaining access to reliable real-time data, and facilitating effective human-machine collaboration under high-stress conditions. Additional complexities arise from addressing issues like bias and fairness in AI models, classifying and categorising systems based on their autonomy and criticality, public perception and ethical concerns, rapid technological advancements, the demand for training and skills development, and the necessity of international collaboration and adherence to evolving regulations and treaties in the ever-evolving landscape of military autonomous and AI-based systems.

RQ4. Standardisation Alignment & RQ5. Case Study Integration: This two research questions are evaluated through the proposal of a certification framework for military avionics (Chapter 4 Methodology) and its validation through its applicability to a military case study, such as visual AAR guidance (Chapter 5 Case study). A highlight of this chapter is the in-depth analysis of a case study of a military system, which provides valuable practical insights derived from real-world AI/ML integration. This case study serves as a critical link between theory and practice, providing concrete examples and considerations for military aviation stakeholders. The applicability of EASA first guidance material on machine learning applications in the military aviation sector is validated. This meets two objectives: to reduce certification efforts, thereby saving time and money, and to participate in the initiative of other aeronautical projects focused on the future harmonisation of standards between the military and civilian domains whenever feasible. Throughout the case study, some military aspects are not covered by civilian guidelines and are identified as gaps; 'civilian guidelines necessary but not sufficient'.

RQ6. Military-Specific Objectives: Chapter 6.2 Proposal of a STANAG presents the gap analysis that emerged during the implementation of the EASA guidance objectives for ML applications [4], assess whether the development of new military-specific objectives is warranted, and if is necessary, offer possible approaches to address them.

7.2 Summary

In conclusion, the rapid advancement of AI algorithms has brought in a new era of possibilities in aviation, both in the civil and military aviation sectors. These algorithms offer the promise of improved situational awareness, faster response times and reduced pilot workload, potentially revolutionising the way aircraft operate. However, these advantages are accompanied by complex challenges related to reliability, accountability, and the ethical use of AI in aviation. While significant efforts have been made in the civil aviation sector to address these challenges, the military avionics domain has received relatively limited attention in this regard.

This study has provided an overview of ongoing efforts to standardise AI technology in civil and military aviation, highlighting areas of alignment and divergence between these sectors. It has also identified and classified a wide range of AI applications in military avionics, highlighting their vital role in improving operational efficiency and mission success. In addition, the research has meticulously analysed the unique certification challenges specific to AI/ML systems in military aviation, delving into issues of safety, adaptability, regulatory compliance, and ethical considerations. The proposal for a military certification framework aligned with ongoing civilian efforts and its validation through a military case study has offered insights and highlighted areas where gaps are identified. The study, at the same time, has explored the alignment of conventional and emerging civilian standards in the military aviation sector, with a focus on harmonisation of standards between the military and civilian domains wherever possible. Finally, the exploration of specific military objectives, including the possible development of a new STANAG, underlines the need for tailored standards to address the complexities of AI-based military systems.

By considering in depth the unique characteristics and operational requirements of military aircraft, this study aims to make a significant contribution to the development of an effective certification framework that facilitates the safe and efficient utilisation of AI/ML systems in this specialised domain.

7.3 Future work

In this study, only a few objectives were assessed for a military case study to address the applicability of the current civilian guidelines. Future work should focus on developing the AI/ML-based model proposed in the case study, training the model, implementing it, and conducting the full certification exercise. Once the full exercise has been completed, a specific STANAG can be developed to address the unique gaps identified in this study with respect to the integration of AI into military avionics. This STANAG should provide comprehensive guidelines and standards specific to military applications, covering aspects such as security assessment, resilience to adverse threats, ethical considerations, and human-AI collaboration. Further activities could link the conclusions of this work with the ongoing activities on AI certification in different fora, e.g. proposing the use case from section 5.1 as a military input for the validation standard ARP6983 in the EUROCAE WG-114/ SAE G-34 scope, developing other use cases relevant for standardisation activities in the EICACS scope, e.g. MALE UAV ISR, and establishing a link with the EDA Trustworthiness for AI in Defence working group. With this in mind, collaboration with military experts, rigorous testing and continuous updates will be essential to ensure the effectiveness and relevance of these military-specific guidelines, ultimately improving the safe and efficient use of AI/ML systems in military aviation.

References

[1] F. E. Morgan *et al.*, "Military Applications of Artificial Intelligence: Ethical Concerns in an Uncertain World," RAND Corporation, Apr. 2020. doi: 10.7249/RR3139-1.

[2] "Artificial Intelligence in Aeronautical Systems: Statement of Concerns - ER-022," EUROCAE, 2021.

[3] C. Torens, U. Durak, and J. Dauer, "Guidelines and Regulatory Framework for Machine Learning in Aviation," in *AIAA SciTech Forum 2022*, American Institute of Aeronautics and Astronautics Inc, AIAA, 2022. doi: 10.2514/6.2022-1132.

[4] "First usable guidance for Level 1 & 2 machine learning applications," *European Aviation Safety Agency*, no. Tech. Rep., 2023.

[5] "Concepts of Design Assurance for Neural Networks (CoDANN)," Tech. rep., European Aviation Safety Agency, 2020.

[6] M. Gray and A. Ertan, "Artificial Intelligence and Autonomy in the Military: An Overview of NATO Member States' Strategies and Deployment," NATO Cooperative Cyber Defence Centre of Excellence, CCDCOE, 2021. [Online]. Available: www.ccdcoe.org

[7] "Perspective on AI," DARPA, U.S. Department of Defense,. Accessed: Sep. 17, 2023. [Online]. Available: https://www.darpa.mil/about-us/darpa-perspective-on-ai

[8] "Artificial Intelligence: Emerging Opportunities, Challenges, and Implications," United States Government Accountability Office, 2018.

[9] "Science and Technology Trends 2020-2040: Exploring the S and T Edge," NATO Science & Technology Organization, 2020.

[10] "Artificial Intelligence Roadmap 2.0: Human-centric approach to AI in aviation Human-centric approach to AI in aviation," European Aviation Safety Agency, 2023.

[11] "Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts," EU Commission, 2021.

[12] S. Kainikara, "Artificial Intelligence and the Future of Air Power," Air Power Development Centre, Royal Australian Air Force, 2019.

[13] "EASA Artificial Intelligence Roadmap 1.0: A human-centric approach to AI in aviation," 2020.

[14] F. Mamalet *et al.*, “White Paper Machine Learning in Certified Systems,” IRT Saint Exupéry, 2021. [Online]. Available: https://hal.archives-ouvertes.fr/hal-03176080

[15] T. Sheridan, “Humans and automation: System design and research issues.,” 2002.

[16] “Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles,” SAE International, 2021.

[17] A. Schweiger *et al.*, “Classification for Avionics Capabilities Enabled by Artificial Intelligence,” in *AIAA/IEEE Digital Avionics Systems Conference*, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/DASC52595.2021.9594364.

[18] M. Tonnin, “Artificial Intelligence: Implications for NATO’s Armed Forces,” Sub-Committee on Technology Trends and Security (STCTTS), STC , 2019.

[19] *Guidelines for Development of Civil Aircraft and Systems*. SAE International SAE Std. ARP4754A, 2012.

[20] “Concepts of Design Assurance for Neural Networks (CoDANN) II,” Tech. rep., European Aviation Safety Agency, 2021.

[21] F. Aviation Agency, “Neural Network Based Runway Landing Guidance for General Aviation Autoland,” 2021.

[22] I. Moir and A. Seabridge, *Military Avionics Systems*. 2006.

[23] K. Biswas and K. Biswas, “Military Aviation Principles,” *Military Engineering*, Sep. 2019, doi: 10.5772/INTECHOPEN.87087.

[24] I. Szabadföldi, “Artificial Intelligence in Military Application – Opportunities and Challenges,” *Land Forces Academy Review*, vol. 26, no. 2, pp. 157–165, Jun. 2021, doi: 10.2478/raft-2021-0022.

[25] “The Responsible Use of Artificial Intelligence in FCAS-An initial Assessment,” Airbus Defence and Space, 2021. doi: 10.13140/RG.2.2.20218.44483.

[26] “Convention on International Civil Aviation,” *Chicago Convention*, 1994.

[27] *Commission Delegated Regulation (EU) 2021/1087*. 2021.

[28] K. Biswas, “Unification of Military and Civil Aircraft Certification Procedures,” *Proceedings of ICTACEM 2014, IIT Kharagpur, India* , Jan. 2015.

[29] O. Hirling and F. Holzapfel, “EASA’s ‘open’ category for military UAS: Opportunities and limitations in the field of airworthiness,” *Aerospace*, vol. 5, no. 3, Sep. 2018, doi: 10.3390/aerospace5030070.

[30] *Commission Regulation (EU) No 748/2012*. European Comission, 2012.

[31] *Software Considerations in Airborne Systems and Equipment Certification*. RTCA, Inc. Std. RTCA DO-178C, 2011.

[32] *Guide and Methods for Safety Assessment*. SAE International SAE Std. ARP4761.

[33] V. Hilderman, “Military Avionics Compliance to DO-178C/254, and ARP4754A/4761/A,” in *Military Certification via DO-178C/254, and ARP4754A/4761/A*, 2021.

[34] M. Dronjak, M. Janković, V. Parezanović, and S. Jovičić, “Military Airworthiness,” in *6th International Scientific Conference on Defensive Techonologies*, 2014. doi: 10.13140/2.1.4391.6806.

[35] “Frequently Asked Questions (FAQs),” Military Airworthiness Authorities (MAWA) Forum, 2015.

[36] “Approved MAWA Documents,” European Defence Agency. Accessed: Sep. 19, 2023. [Online]. Available: https://eda.europa.eu/experts/airworthiness/mawa-documents

[37] “Press release EASA and EDA strengthen civil-military cooperation in aviation safety,” 2013.

[38] LufABw, “Zulassung von Produkten, Bau- und Ausrüstungsteilen sowie Genehmigung von Entwicklern und Herstellern - DEMAR 21.”

[39] “Standardization,” NATO. Accessed: Sep. 19, 2023. [Online]. Available: https://www.nato.int/cps/en/natohq/topics_69269.htm

[40] *Modular and Open Avionics Architectures*. NATO STANAG-462, 2005.

[41] *Guidance on Software Safety Design and Assessment of Munition-related Computing Systems*. NATO STANAG 4452, 2016.

[42] *A Defense System Software Development*. DOD-STD-2167, 1994. Accessed: Sep. 19, 2023. [Online]. Available: http://everyspec.com/DoD/DoD-STD/DOD-STD-2167A_8470/

[43] *Software Development and Documentation*. MIL-STD-498, 1994. Accessed: Sep. 19, 2023. [Online]. Available: http://everyspec.com/MIL-STD/MIL-STD-0300-0499/MIL-STD-498_25500/

[44] *System Safety*. MIL-STD-882, 2012. Accessed: Sep. 19, 2023. [Online]. Available: http://everyspec.com/MIL-STD/MIL-STD-0800-0899/MIL-STD-882E_41682/

[45] *A400M-Military Certification Review Item (MCRI)*. MCRI F-06.

[46] “A400M Programme EASA and CQO Collaboration Lessons learnt from the joint civil-military certification approach on a military transport aircraft”.

[47] A. Monzón Diaz, "Towards a regulatory framework to certify artificial intelligence applications in aviation," 2023.

[48] "EASA Concept Paper: First usable guidance for Level 1 machine learning applications," *European Aviation Safety Agency*, no. Tech. Rep, 2021.

[49] J. Marques and A. M. Da Cunha, "A Set of Requirements for Certification of Airborne Military Software," *AIAA/IEEE Digital Avionics Systems Conference - Proceedings*, vol. 2019-September, Sep. 2019, doi: 10.1109/DASC43569.2019.9081775.

[50] "An AI Just Beat a Human F-16 Pilot In a Dogfight — Again - Defense One." Accessed: Sep. 20, 2023. [Online]. Available: https://www.defenseone.com/technology/2020/08/ai-just-beat-human-f-16-pilot-dogfight-again/167872/

[51] M. Damour *et al.*, "Towards Certification of a Reduced Footprint ACAS-Xu System: a Hybrid ML-based Solution," p. 978, 2021, doi: 10.1007/978-3-030-83903-1_3ï.

[52] C. Gabreau *et al.*, "Toward the certification of safety-related systems using ML techniques: the ACAS-Xu experience," 2022, [Online]. Available: https://hal.science/hal-03761946

[53] Daedalean, "Situational Intelligence™ in the skies Bringing people closer through accessible air mobility enabled by AI technology."

[54] Daedalean, "The Future of Avionics: High Performance, Machine Learned and Certified."

[55] "Overview — Multi-Color Infrared Alerting Sensor (MIRAS) — Electronic Support Measures/Electronic Warfare — Sensors/Electronics — Weapons — Military Periscope." Accessed: Oct. 30, 2023. [Online]. Available: https://www.militaryperiscope.com/weapons/sensorselectronics/electronic-support-measureselectronic-warfare/multi-color-infrared/overview/

[56] K. Dmitriev, J. Schumann, and F. Holzapfel, "Toward Certification of Machine-Learning Systems for Low Criticality Airborne Applications," in *AIAA/IEEE Digital Avionics Systems Conference - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/DASC52595.2021.9594467.

[57] K. Dmitriev, J. Schumann, and F. Holzapfel, "Toward Design Assurance of Machine-Learning Airborne Systems," in *AIAA Science and Technology Forum and Exposition, AIAA SciTech Forum 2022*, American Institute of Aeronautics and Astronautics Inc, AIAA, 2022. doi: 10.2514/6.2022-1134.

[58] K. Dmitriev, J. Schumann, and F. Holzapfel, "Towards Design Assurance Level C for Machine-Learning Airborne Applications," in *AIAA/IEEE Digital Avionics Systems Conference - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/DASC55683.2022.9925741.

[59] “SC-AI-01 Trustworthiness of Machine Learning based Systems,” European Aviation Safety Agency, 2022.

[60] MLEAP Consortium, “EASA Research – Machine Learning Application Approval (MLEAP) interim technical report,” European Union Aviation Safety Agency, 2023.

[61] “Ethics guidelines for trustworthy AI | Shaping Europe’s digital future.” Accessed: Sep. 20, 2023. [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[62] “Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment”, doi: 10.2759/791819.

[63] “MEPs ready to negotiate first-ever rules for safe and transparent AI | News | European Parliament.” Accessed: Sep. 20, 2023. [Online]. Available: https://www.europarl.europa.eu/news/en/press-room/20230609IPR96212/meps-ready-to-negotiate-first-ever-rules-for-safe-and-transparent-ai

[64] “CS-23 Amendment 6 and AMC & GM to CS-23 Issue 4 - Normal-Category Aeroplanes (CS-23) | EASA.” Accessed: Sep. 20, 2023. [Online]. Available: https://www.easa.europa.eu/en/document-library/certification-specifications/cs-23-amendment-6-and-amc-gm-cs-23-issue-4

[65] M. consortium *et al.*, “MLEAP STAKEHOLDERS DAY 2 - 17 May 2023 - EASA Day 2.”

[66] “ARP6983 (WIP) Process Standard for Development and Certification/Approval of Aeronautical Safety-Related Products Implementing AI - SAE International.” Accessed: Sep. 20, 2023. [Online]. Available: https://www.sae.org/standards/content/arp6983/

[67] “AIR6987 (WIP) Artificial Intelligence in Aeronautical Systems: Taxonomy - SAE International.” Accessed: Sep. 20, 2023. [Online]. Available: https://www.sae.org/standards/content/air6987/

[68] “F3269 Standard Practice for Methods to Safely Bound Behavior of Aircraft Systems Containing Complex Functions Using Run-Time Assurance.” Accessed: Sep. 20, 2023. [Online]. Available: https://www.astm.org/f3269-21.html

[69] “TOVEK | News.” Accessed: Sep. 20, 2023. [Online]. Available: https://www.tovek.cz/en/news.html

[70] “EICACS | European Defence Fund.” Accessed: Sep. 20, 2023. [Online]. Available: https://www.eicacs.eu/

[71] “Home - AI for Defence.” Accessed: Sep. 20, 2023. [Online]. Available: https://ai4def.com/

[72] “Defense Applications of AI - AIAI 2022 Conference.” Accessed: Sep. 20, 2023. [Online]. Available: https://ifipaiai.org/2022/workshops/daai/

[73] “AI Action Plan Workshop: Trustworthiness and Standardisation for AI in Defence.” Accessed: Sep. 20, 2023. [Online]. Available: https://eda.europa.eu/news-and-events/events/2023/09/05/default-calendar/ai-action-plan-workshop-trustworthiness-and-standardisation-for-ai-in-defence

[74] “Projekt »KIEZ 4.0« - Fraunhofer IAIS.” Accessed: Oct. 30, 2023. [Online]. Available: https://www.iais.fraunhofer.de/de/forschung/kuenstliche-intelligenz/projekt--kiez-4-0-.html

[75] Z. M. Qian A Jabbar Xiaolong Li Editors, “Lecture Notes in Electrical Engineering 942 Proceeding of 2021 International Conference on Wireless Communications, Networking and Applications.” [Online]. Available: https://link.springer.com/bookseries/7818

[76] “In-Flight Refueling Video Systems | Kappa optronics: Camera & Vision Systems.” Accessed: Sep. 25, 2023. [Online]. Available: https://www.kappa-optronics.com/en/aviation/in-flight-refueling-vision-systems/

[77] “Artificial Intelligence is the Future of Warfare (Just Not in the Way You Think) - Modern War Institute.” Accessed: Sep. 28, 2023. [Online]. Available: https://mwi.usma.edu/artificial-intelligence-future-warfare-just-not-way-think/

[78] “A-9-2021-0001_EN,” European Parliment, 2021.

[79] F. Sauer, “Artificial Intelligence in the Armed Forces On the need for regulation regarding autonomy in weapon systems”.

[80] “Multinational MRTT fleet standards related document,” 2023. [Online]. Available: http://www.japcc.org/aar

Appendix A: Classification of military AI-based airborne systems

The table categorizes these applications according to their operational domains. It is important to note that the focus here is on the different systems that include AI, rather than platforms. A single platform can possess multiple systems that perform various operations simultaneously. For example, UAVs (platform) can perform autonomous flight (autonomous flight system), flight formation, surveillance, and reconnaissance (ISR system), and so on. Each military application domain is associated with avionics subdomain, predominant AI techniques employed, benefits realized in the defence domain, autonomy level allocation and a criticality assessment:

- Military application domains: specific military-related areas or sectors where these AI-based systems are applied to enhance or support various functions and operations.
- Avionics subdomains: Avionics subdomains that can be found in the main military application domains.
- Benefits: This criterion focuses on the advantages and benefits that AI/ML integration brings to individual military systems. It encompasses improvements in efficiency, accuracy, reduced human workload, enhanced situational awareness, and other system-specific advantages.
- AI technique: This classification considers all AI techniques employed in military systems, not limited to ML, and supervised learning. It aims to highlight the diverse range of AI types used following the taxonomy of EASA Roadmap high kited in Figure 2-2:
    1. Machine learning (ML) with supervised learning, unsupervised learning, or reinforcement learning (RL). Deep learning (DL) which include the NNs. These neuronal networks can be applied to natural language processing (NPL) through RNNs or to computer vision (CV) through CNNs.
    2. Logic and knowledge-based approaches (LKB): fuzzy logic, expert systems, etc.
    3. Statistical approaches (SA): Bayesian estimation, search, and optimisation methods.
- Criticality: It distinguishes between systems that are safety critical solely in a military context (indicated as 'critical (M)'), and those that are safety critical in both civil and military domains (indicated as 'critical').

Applications Domain	AI-based military airborne sub-domain	Benefits	AI technique	Criticality
Autonomous Flight Systems	Flight formation	Reduced risk to collision	ML (RL)	Critical **(M)**
	Flight control	Access to hard-reachable areas, enhanced manoeuvrability	ML (DL)	Critical
	Autonomous navigation	Improved flight efficiency and safety	ML (DL)	Critical
	Autopilot	Reduction pilot workload	ML (DL)	Critical
	Autonomous taking off and landing	Enhanced mission execution	ML (CV)	Critical
	Collision avoidance	Rapid response and evasive manoeuvres	ML (RL)	Critical **(M)**
Intelligence, Surveillance and Reconnaissance (ISR)	Threat detection	Enhanced situational awareness Faster and more efficient surveillance and reconnaissance	ML	Non-critical
	Pattern recognition	Battlespace Awareness	SA & ML	Non-critical
	Situational Awareness (SA)	Fast processing of vast amounts of data	ML (CV & NPL)	Critical **(M)**
Cyber Security and Resilience (CSR)	Resilient autonomy	Resilience against cyber-attacks, safeguarding sensitive military information	LKB	Critical
	Cyber warfare	Reconnaissance, vulnerability scanning, and even exploit selection	ML (CV)	Critical **(M)**
Mission Planning and Execution (MPE)	Dynamic route planning	Real-time adaptability	ML (RL)	Non-critical
	Mission optimization	Improved mission success rates, streamlined mission planning and execution, optimal resource allocation and coordination	SA	Non-critical
	Fuel efficiency	Cost reduction	SA	Non-critical
Countermeasures	Autonomous defence aids systems (DASS)	Rapid response to threats	LKB	Critical **(M)**
	Electronic warfare (EW)	Vulnerability scanning and adaptive responses	ML & LKB	Non-critical
Autonomous Weapon Capability	Battle damage assessment	Enhanced situational awareness	ML (CV)	Non-critical
	Threat Assessment and Aiming Analysis (TAA)	Precision and accuracy	LKB	Non-critical
	Weapon selection & trajectory planning	Accuracy, convenience, and precision	ML (RL) & LKB	Non-critical **(M)**
Decision Support Systems	Speech	Fast complex decision-making	ML (NPL)	Critical
	Visual guidance	Enhanced precision, real-time object recognition and reduced workload	ML (CV)	
Command, Control, Computers, Communications, and Intelligence (C4I)	Interoperability	Increased productivity, mission efficiency, optimal resources allocation and coordination	ML (NLP) & data fusion	Critical **(M)**
	Data fusion and intelligence analyses	Battlespace Awareness, Information Dominance, Decision Advantage	ML(DL) & data fusion	Critical **(M)**
Training and Simulation	Pilot and mission operator training	Improved and customized training and customized training environments, cost effective	ML	Non-critical
	Combat and flight simulation	Refining tactics and strategies	ML (DL)	Non-critical
Air Traffic Management (ATM)	Traffic flow optimization	Enhanced mission efficiency and prioritization	SA	Non-critical
	Fleet management	Optimal resource allocation and coordination	ML & SA	Non-critical
Predictive Maintenance	In-flight system monitoring	Early fault detection	ML & LKB	Non-critical
	Real-time fault diagnosis	Minimize system failure	ML & LKB	Non-critical
	Maintenance forecasting	Labour and cost reduction and improve inventory and repairs management	ML & SA	Non-critical

Table 7-1: Classification of military AI-based airborne systems

Appendix B: EASA objectives for AI level 1 & 2

EASA building blocks	Sub-building blocks	Example	DAL A SWAL 1	DAL B -	DAL C SWAL 2	- SWAL 3	DAL D SWAL 4
Trustworthiness analysis	Characterization of AI	**CO-01:** The applicant should identify the list of end-users that are intended to interact with the AI-based system, together with their roles, their responsibilities, and their expected expertise (including assumptions made on the level of training, qualification, and skills).	○	○	○	○	○
		CO-02: For each end-user, the applicant should identify which high-level task(s) are intended to be performed in interaction with the AI-based system.	○	○	○	○	○
		CO-03: The applicant should determine the AI-based system taking into account domain-specific definitions of 'system'.	○	○	○	○	○
		CO-04: The applicant should define and document the ConOps for the AI-based system, including the task allocation pattern between the end-user(s) and the AI- based system. A focus should be put on the definition of the OD and on the capture of specific operational limitations and assumptions.	○	○	○	○	○
		CO-05: The applicant should document how end-users' inputs are collected and accounted for in the development of the AI-based system.	○	○	○	○	○
		CO-06: The applicant should perform a functional analysis of the system.	○	○	○	○	○
		CL-01: The applicant should classify the AI-based system, based on the levels presented in Table 2, with adequate justifications.	○	○	○	○	○
	Safety Assessment	**SA-01:** The applicant should perform a safety (support) assessment for all AI-based (sub)systems, identifying and addressing specificities introduced by AI/ML usage.	●	●	○	○	○
		ICSA-01: The applicant should identify which data needs to be recorded for the purpose of supporting the continuous safety assessment.	●	●	○	○	○
		ICSA-02: The applicant should use the collected data to perform a continuous safety assessment. This includes: — the definition of target values, thresholds and evaluation periods to guarantee that design assumptions hold; — the monitoring of in-service events to detect potential issues or suboptimal performance trends that might contribute to safety margin erosion, or, for non-ATS providers, to service performance degradations; and — the resolution of identified shortcomings or issues.	●	●	○	○	○
		IS-01: For each AI-based system and its data sets, the applicant should identify those information security risks with an impact on safety, identifying and addressing specific threats introduced by AI/ML usage.	○	○	○	○	○
	Information Security Assessment	**IS-01:** For each AI-based system and its data sets, the applicant should identify those information security risks with an impact on safety, identifying and addressing specific threats introduced by AI/ML usage.	○	○	○	○	○
		IS-02: The applicant should document a mitigation approach to address the identified AI/ML-specific security risk. Note: Beyond the applicability defined here for any domain, further levelling may be introduced in domains defining specific security assurance levels (SALs). See Section D.3.	○	○	○	○	○
		IS-03: The applicant should validate and verify the effectiveness of the security controls introduced to mitigate the identified AI/ML-specific security risks to an acceptable level.	○	○	○	○	○

		Note: Beyond the applicability defined here for any domain, further levelling may be introduced in domains defining specific security assurance levels (SALs). See Section D.3.					
	Ethics-based Assessment	**ET-01:** The applicant should perform an ethics-based trustworthiness assessment for any AI-based system developed using ML techniques or incorporating ML models.	○	○	○	○	○
		ET-02: The applicant should ensure that the AI-based system bears no risk of creating human attachment, stimulating addictive behaviour, or manipulating the end-user's behaviour.	○	○	○	○	○
		ET-03: The applicant should ensure that the AI-based system presents no capability of adaptive learning.	○	○	○	○	○
		ET-04: The applicant should comply with national and EU data protection regulations (e.g. GDPR), i.e. involve their Data Protection Officer (DPO), consult with their National Data Protection Authority, etc.	○	○	○	○	○
		ET-05: The applicant should ensure that procedures are in place to avoid creating or reinforcing unfair bias in the AI-based system, regarding both the data sets and the trained models.	○	○	○	○	○
		ET-06: The applicant should perform an environmental impact analysis, identifying and assessing potential negative impacts of the AI-based system on the environment and human health throughout its life cycle (development, deployment, use, end of life).	○	○	○	○	○
		ET-07: The applicant should define measures to reduce or mitigate the impacts identified under Objective ET-06.	○	○	○	○	○
		ET-08: The applicant should identify the need for new competencies for users and end-users to interact with and operate the AI-based system and mitigate possible training gaps (link to **Provision ORG-06, Provision ORG- 07**).	○	○	○	○	○
		ET-09: The applicant should perform an assessment of the risk of de-skilling of the users and end-users and mitigate the identified risk through a training needs analysis and a consequent training activity (link to **Provision ORG-06, Provision ORG-07**).	○	○	○	○	○
AI Assurance	Learning Assurance	**DA-01:** The applicant should describe the proposed learning assurance process, taking into account each of the steps described in Sections C.3.1.2 to C.3.1.12, as well as the interface and compatibility with development assurance processes.	○	○	○	○	○
		DA-02: Documents should be prepared to encompass the capture of the following minimum requirements: — safety requirements allocated to the AI/ML constituent; — information security requirements allocated to the AI/ML constituent; — functional requirements allocated to the AI/ML constituent; — operational requirements allocated to the AI/ML constituent, including ODD and AI/ML constituent performance monitoring, detection of OoD input data and data-recording requirements; — non-functional requirements allocated to the AI/ML constituent (e.g. performance, scalability, reliability, resilience, etc.); and — interface requirements.	○	○	○	○	○
		DA-03: The applicant should describe the system and subsystem architecture, to serve as reference for related safety (support) assessment and learning assurance objectives.	○	○	○	○	
		DA-04: Each of the captured requirements should be validated.	●	●	○	○	○
		DA-05: The applicant should document evidence that all derived requirements have been provided to the (sub)system processes, including the safety (support) assessment.	○	○	○	○	○
		DA-06: The applicant should document evidence of the validation of the derived requirements, and of the determination of any impact on the safety (support) assessment and (sub)system requirements.	○	○	○	○	○

		DA-07: Each of the captured (sub)system requirements allocated to the AI/ML constituent should be verified.	●	●	○	○	○
		DM-01: The applicant should define the set of parameters pertaining to the AI/ML constituent ODD.	○	○	○	○	○
		DM-02: The applicant should capture the DQRs for all data pertaining to the data management process, including but not limited to: — the data needed to support the intended use; — the ability to determine the origin of the data; — the requirements related to the annotation process; — the format, accuracy and resolution of the data; — the traceability of the data from their origin to their final operation through the whole pipeline of operations; — the mechanisms ensuring that the data will not be corrupted while stored or processed, — the completeness and representativeness of the data sets; and — the level of independence between the training, validation and test data sets.	○	○	○	○	○
		DM-03: The applicant should capture the requirements on data to be pre-processed and engineered for the inference model in development and for the operations.	○	○	○	○	○
		DM-04: The applicant should ensure the validation to an adequate level of the correctness and completeness of the ML constituent ODD.	●	●	○	○	○
		DM-05: The applicant should ensure the validation of the correctness and completeness of requirements on data to be pre-processed and engineered for the trained and inference model, as well as of the DQRs on data.	●	●	○	○	○
		DM-06: The applicant should identify data sources and collect data in accordance with the defined ODD, while ensuring satisfaction of the defined DQRs, in order to drive the selection of the training, validation and test data sets.	○	○	○	○	○
		DM-07: Once data sources are collected, the applicant should ensure the high quality of the annotated or labelled data in the data set.	●	●	○	○	○
		DM-08: The applicant should define the data preparation operations to properly address the captured requirements (including DQRs).	○	○	○	○	○
		DM-09: The applicant should define and document pre- processing operations on the collected data in preparation of the training.	○	○	○		
		DM-10: When applicable, the applicant should define and document the transformations to the pre-processed data from the specified input space into features which are effective for the performance of the selected learning algorithm.	○	○	○		
		DM-11: If the learning algorithm is sensitive to the scale of the input data, the applicant should ensure that the data is effective for the stability of the learning process.	○	○	○		
		DM-12: The applicant should distribute the data into three separate and independent data sets which will meet the specified DQRs: — the training data set and validation data set, used during the model training; — the test data set used during the learning process verification, and the inference model verification.	○	○	○	○	○
		DM-13: The applicant should ensure validation and verification of the data, as appropriate, all along the data management process so that the data management requirements (including the DQRs) are addressed.	●	●	○	○	○
		DM-14: The applicant should perform a data and learning verification step to confirm that the appropriate data sets have been used for the training, validation and verification of the model and that the expected guarantees (generalisation, robustness) on the model have been reached.	●	●			
		LM-01: The applicant should describe the AI/ML constituents and the model architecture.	○	○	○	○	○

		LM-02: The applicant should capture the requirements pertaining to the learning management and training processes, including but not limited to: — model family and model selection; — learning algorithm(s) selection; — cost/loss function selection describing the link to the performance metrics; — model bias and variance metrics and acceptable levels; — model robustness and stability metrics and acceptable levels; — training environment (hardware and software) identification; — model parameters initialisation strategy; — hyper-parameters identification and setting; — expected performance with training, validation and test sets.	○	○	○	○	○
		LM-03: The applicant should document the credit sought from the training environment and qualify the environment accordingly.	○	○	○		
		LM-04: The applicant should provide quantifiable generalisation guarantees.	○	○	○		
		LM-05: The applicant should document the result of the model training.	○	○	○	○	○
		LM-06: The applicant should document any model optimisation that may affect the model behaviour (e.g. pruning, quantisation) and assess their impact on the model behaviour or performance.	○	○	○		
		LM-07: The applicant should account for the bias- variance trade-off in the model family selection and should provide evidence of the reproducibility of the training process.	●	●	○		
		LM-08: The applicant should ensure that the estimated bias and variance of the selected model meet the associated learning process management requirements.	●	●	○	○	○
		LM-09: The applicant should perform an evaluation of the performance of the trained model based on the test data set and document the result of the model verification.	●	●	○	○	○
		LM-10: The applicant should perform a requirements- based verification of the trained model behaviour and document the coverage of the AI/ML constituent requirements by verification methods.	●	●	○	○	○
		LM-11: The applicant should provide an analysis on the stability of the learning algorithms.	●	●	○		
		LM-12: The applicant should perform and document the verification of the stability of the trained model.	●	●	○	○	○
		LM-13: The applicant should perform and document the verification of the robustness of the trained model in adverse conditions.	●	●	○	○	○
		LM-14: The applicant should verify the anticipated generalisation bounds using the test data set.	●	●	○		
		IMP-01: The applicant should capture the requirements pertaining to the implementation process.	○	○	○	○	○
		IMP-02: Any post-training model transformation (conversion, optimisation) should be identified and validated for its impact on the model behaviour and performance, and the environment (i.e. software tools and hardware) necessary to perform model transformation should be identified.	○	○	○		
		IMP-03: The applicant should plan and execute appropriate development assurance processes to develop the inference model into software and/or hardware items.	○	○	○		
		IMP-04: The applicant should verify that any transformation (conversion, optimisation, inference model development) performed during the trained model implementation step has not adversely altered the defined model properties.	○	○	○		
		IMP-05: The differences between the software and hardware of the platform used for training and the one used for verification should be identified and assessed for their possible impact on the inference model behaviour and performance.	○	○	○		

		MP-06: The applicant should perform an evaluation of the performance of the inference model based on the test data set and document the result of the model verification.	○	○	○	○	○
		IMP-07: The applicant should perform and document the verification of the stability of the inference model.	●	●	○	○	○
		IMP-08: The applicant should perform and document the verification of the robustness of the inference model in adverse conditions.	●	●	○	○	○
		IMP-09: The applicant should perform a requirements- based verification of the inference model behaviour when integrated into the AI/ML constituent and document the coverage of the ML constituent requirements by verification methods.	●	●	○	○	○
		CM-01: The applicant should apply all configuration management principles to the AI/ML constituent life- cycle data, including but not limited to: — identification of configuration items; versioning; baselining; — change control; reproducibility; — problem reporting; — archiving and retrieval, and retention period.	●	●	●	●	●
		QA-01: The applicant should ensure that quality/process assurance principles are applied to the development of the AI-based system, with the required independence level.	○	○	○	○	○
	Development Explainability	**EXP-01:** The applicant should identify the list of stakeholders, other than end-users, that need explainability of the AI-based system at any stage of its life cycle, together with their roles, their responsibilities and their expected expertise (including assumptions made on the level of training, qualification and skills).	○	○	○		
		EXP-02: For each of these stakeholders (or groups of stakeholders), the applicant should characterise the need for explainability to be provided, which is necessary to support the development and learning assurance processes.	○	○	○		
		EXP-03: The applicant should identify and document the methods at AI/ML item and/or output level satisfying the specified AI explainability needs.	○	○	○		
		EXP-04: The applicant should provide the means to record operational data that is necessary to explain, post operations, the behaviour of the AI-based system and its interactions with the end-user.	○	○	○	○	○
Human Factors for AI	Operational Explainability	**EXP-05:** For each output of the AI-based system relevant to task(s) (per Objective CO-02), the applicant should characterise the need for explainability.	○	○	○	○	○
		EXP-06: The applicant should present explanations to the end-user in a clear and unambiguous form.	○	○	○	○	○
		EXP-07: The applicant should define relevant explainability so that the receiver of the information can use the explanation to assess the appropriateness of the decision / action as expected.	○	○	○	○	○
		abstraction of the explanations, taking into account the characteristics of the task, the situation, the level of expertise of the end-user and the general trust given to the system.	○	○	○	○	○
		EXP-08: The applicant should define the level of	○	○	○	○	○
		EXP-09: Where a customisation capability is available, the end-user should be able to customise the level of details provided by the system as part of the explainability.	○	○	○	○	○
		EXP-10: The applicant should define the timing when the explainability will be available to the end-user taking into account the time criticality of the situation, the needs of the end-user, and the operational impact.	○	○	○	○	○
		EXP-11: The applicant should design the AI-based system so as to enable the end-user to get upon request explanation or additional details on the explanation when needed.	○	○	○	○	○

		EXP-12: For each output relevant to the task(s), the applicant should ensure the validity of the specified explanation, based on actual measurements (e.g. monitoring) or on a quantification of the level of uncertainty.	O	O	O	O	O
		EXP-14: The AI-based system inputs should be monitored to be within the operational boundaries (both in terms of input parameter range and distribution) in which the AI/ML constituent performance is guaranteed, and deviations should be indicated to the relevant users and end-users.	O	O	O	O	O
		EXP-15: The AI-based system outputs should be monitored to be within the specified operational performance boundaries, and deviations should be indicated to the relevant users and end-users.	O	O	O	O	O
		EXP-16: The training and instructions available for the human end-user should include procedures for handling possible outputs of the ODD and performance monitoring.	O	O	O	O	O
		EXP-17: Information concerning unsafe AI-based system operating conditions should be provided to the human end-user to enable them to take appropriate corrective action in a timely manner.	O	O	O		
	Human-AI teaming	**HF-01:** The applicant should design the AI-based system with the ability to build its own individual situational awareness.	O	O	O	O	O
		HF-02: The applicant should design the AI-based system with the ability to allow the end-user to ask questions and to answer questions from the end-user, in order to reinforce the end-user individual situational awareness.	O	O	O	O	O
		HF-03: The applicant should design the AI-based system with the ability to modify its individual situational awareness on end-user request.	O	O	O	O	O
		HF-04: If a decision is taken by the AI-based system, the applicant should design the AI-based system with the ability to request from the end-user a cross-check validation. Corollary objective: The applicant should design the AI-based system with the ability to cross-check and validate a decision made by the end-user automatically or on request.	O	O	O	O	O
		HF-05: For complex situations under normal operations, the applicant should design the AI-based system with the ability to identify suboptimal strategy and propose through argumentation an optimised solution. Corollary objective: The applicant should design the AI- based system with the ability to accept rejection required by the end-user on the proposal.	O	O	O	O	O
		HF-06: For complex situations under abnormal operations, the applicant should design the AI-based system with the ability to identify the problem, share the diagnosis including the root cause, the resolution strategy and the anticipated operational consequences. Corollary objective: The applicant should design the AI- based system with the ability to consider the arguments shared by the end-user.	O	O	O	O	O
		HF-07: The applicant should design the AI-based system with the ability to detect poor decision-making by the end-user in a time-critical situation.	O	O	O	O	O
		HF-08: The applicant should design the AI-based system with the ability to take the appropriate action outside of a collaboration scheme, in case of detection of poor decision-making by the end-user in a time-critical situation.	O	O	O	O	O
		HF-09: The applicant should design the AI-based system with the ability to negotiate, argue, and support its positions.	O	O	O	O	O
		HF-10: The applicant should design the AI-based system with the ability to accept the modification of task allocation / task adjustments (instantaneous/short- term).	O	O	O	O	O

	Modality of Interaction	**HF-11:** The applicant should design the AI-based system with the ability to understand through the end-user responses or his or her action that there was a misinterpretation from the end-user.	O	O	O	O	O
		HF-12: The applicant should design the AI-based system with the ability to notify the end-user that he or she misunderstood the information provided through spoken natural language.	O	O	O	O	O
		HF-13: In case of misinterpretation, the applicant should design the AI-based system with the ability to resolve the misunderstanding through repetition, modification of the modality of interaction and/or with the provision of additional rationale on the initial information.	O	O	O	O	O
		HF-14: The applicant should design the AI-based system with the ability to making room for dialogue turns by other participants, keeping silent when needed and not hindering the user (while user is engaged in a complex activity), sticking to information that answers a given question by other participants, etc.).	O	O	O	O	O
		HF-15: In case spoken natural language is used, the applicant should design the AI-based system with the ability to provide information regarding the associated AI- based system capabilities and limitations.	O	O	O	O	O
		HF-16: The applicant should design the syntax of the spoken procedural language so that it can be learned easily by the end-user.	O	O	O	O	O
		HF-17: The applicant should design the AI-based system with the ability to transition from verbal natural language to verbal procedural language depending on its own perception of the performance of the dialogue, the context of the situation and the characteristics of the task.	O	O	O	O	O
		HF-18: The applicant should design the gesture language syntax so that they are intuitively associated with the command that they are supposed to trigger.	O	O	O	O	O
		HF-19: The applicant should design the AI-based system with the ability to filter the intentional gesture used for language from non-intentional gesture, such as spontaneous gestures that are made to complement verbal spoken language.	O	O	O	O	O
		HF-20: In case gesture non-verbal language is used, the applicant should design the AI-based system with the ability to recognise the end-user intention.	O	O	O	O	O
		HF-21: In case gesture non-verbal language is used, the applicant should design the AI-based system with the ability to acknowledge the end-user intention with appropriate feedback.	O	O	O	O	O
		HF-22: In case spoken natural language is used, the applicant should design the AI-based system so that it can be deactivated for the benefit of other modalities in case of degradation of the interaction performance.	O	O	O	O	O
		HF-23: The applicant should design the AI-based system with the ability to combine or adapt the interaction modalities depending on the characteristics of the task, the operational event, and/or the operational environment.	O	O	O	O	O
		HF-24: The applicant should design the AI-based system with the ability to automatically adapt the modality of interactions to the end-user states, the situation, the context and/or the perceived end-user's preferences.	O	O	O	O	O
		HF-25: The applicant should design the AI-based system to minimise the likelihood of design related error made by the end-user.	O	O	O	O	O
		HF-26: The applicant should design the AI-based system to minimise the likelihood of design related error made by the Human-AI Teaming.	O	O	O	O	O
		HF-27: The applicant should design the AI-based system to minimise the likelihood of HAIRM related error.	O	O	O	O	O

		HF-28: The applicant should design the AI-based system to be tolerant to end-user error.	○	○	○	○	○
		HF-29: The applicant should design the AI-based system so that in case the end-user makes an error while interacting with AI-based system, the opportunities exist to detect the error.	○	○	○	○	○
		HF-30: The applicant should design the AI-based system so that once an error is detected, the AI-based system should provide efficient means to inform the end-user and correct the error.	○	○	○	○	○
AI Safety Risk Mitigation	AI Safety Risk Mitigation	**SRM-01:** Once activities associated with all other building blocks are defined, the applicant should determine whether the coverage of the objectives associated with the explainability and learning assurance building blocks is sufficient or if an additional dedicated layer of protection, called hereafter safety risk mitigation (SRM), would be necessary to mitigate the residual risks to an acceptable level.	●	●	○	○	
		SRM-02: The applicant should establish SRM means as identified in Objective SRM-01.	●	●	○	○	

Table 7-2: EASA objectives for AI level 1 & 2