OVERVIEW OF INITIATIVES SUITABLE FOR LEARNING ASSURANCE OF AI-BASED MILITARY PRODUCTS AND IDENTIFIED CHALLENGES BASED ON USE CASES ANALYSIS

A. de Cacqueray[1], J. Ribas de Amaral[2], C. M. Capdevila Llompart[1]
[1] Airbus Defence and Space GmbH, Rechliner Straße, 85077 Manching, Germany
[2] Airbus Defence and Space S.A.U, C/ Aviocar, 2, 28906 Getafe, Madrid, Spain

Abstract

Machine learning (ML) can enhance military system performance and safety, crucial for future competitiveness. However, this shift from rule-based to data-driven approaches is not covered by current regulations and standards. To ensure reliability and safety, EASA and EUROCAE WG-114 introduced a learning assurance process for civil aeronautical ML-based systems. This paper explores challenges and specificities of applying the proposed learning assurance process for certifying and qualifying military ML-based systems using relevant use cases, focusing on the requirements identification and architecture definition.

1. INTRODUCTION

The defence sector is characterised by its complex missions, threatening environments, need for interoperability, improved situational awareness, quick analysis of large amounts of data, and fast decision-making. In this context, Artificial Intelligence (AI)-based systems have the potential to transform these military operations and provide advanced capabilities. However, ML models exhibit a probabilistic nature (data-driven), inferring the most likely outcomes for new data based on relationships learned from training datasets. This results in extra difficulties to the certification and qualification process of ML-based systems, as they do not behave or cannot be characterized like traditional systems.

The civil aviation industry, academia and regulatory bodies have started working on guidelines to standardise the learning assurance process of aeronautical products that implement ML. When it comes to military aviation, adapting civil standards is a common practice. However, these standards need to be adapted to the specificities of the military context.

Following the work of [1], which examined the unique aspects of certifying ML systems for military applications considering the European Union Aviation Safety Agency (EASA)'s Trustworthy AI building blocks [2], this paper investigates whether the ongoing standardization efforts, particularly covering the AI Assurance building block, can be applied to develop military ML-based systems that operate safely and effectively. This paper focuses solely on AI learning assurance aspects and does not address information security, ethics, or human factors.

The structure of the paper is the following: first an overview of the existing initiatives of ML regulations and standards, both on the civil and military side, is provided. Then, the learning assurance concept, with the W-shaped process, is presented. Finally, two hypothetical military use cases are presented to illustrate how this process can be applied: enhanced automatic aerial refuelling and automatic target recognition (ATR). These examples are intended to demonstrate the process in action and may not represent real development. They are provided for illustrative purposes only.

2. STATE OF THE ART: OVERVIEW OF EXISTING INITIATIVES OF ML REGULATIONS AND STANDARDS

2.1. Civil

The EU published the AI Act [3] in July 2024, making it the world's first binding regulation on AI. The EU AI Act proposes a comprehensive framework for AI governance, encompassing risk assessment, unacceptable AI practices, transparency, data quality, and human oversight. On the aviation side, the (EASA) is currently working to integrate the requirements outlined in Title III, Chapter 2 of the EU AI Act into future airworthiness, Air Traffic Management/Air Navigation Services (ATM/ANS), and unmanned aircraft regulations. As part of the initial milestones of EASA's AI Roadmap 2.0 Phase I [2], EASA AI Concept Papers [4] [5] were built to explore the developments of early guidance considering full compatibility with the EU AI Act. They outline preliminary objectives for Level 1 AI (human assistance) and Level 2 AI (human-AI collaboration). Its purpose is to support applicants in integrating AI and ML technologies into systems used for safety- or environment-related applications across all domains governed by the EASA Basic Regulation (Regulation (EU) 2018/1139) [6]. Issue 2 of the EASA AI Concept Paper [5] marks the entry of the EASA AI Roadmap into its second phase (framework consolidation), where Rulemaking Task (RMT).0742 [7] will facilitate the integration of the anticipated guidance from the AI Concept Paper into a comprehensive framework of generic rules and acceptable means of compliance (AMC). The initiative focuses on 2 main steps: (1) proposing a regulatory framework for AI trustworthiness in aviation and developing generic AMC and guidance material (GM), and (2) adapting existing regulatory materials to suit both the aviation domains listed in the AI Act and other impacted sectors. As mentioned during EASA AI Days 2025, the first step is planned to focus on soft law, such as detailed specifications (DS), rather than hard law. On the industry side, EUROCAE WG-114 and SAE G-34 are collaborating to develop standards and guidance for the development and certification of AI-based airborne and

doi: 10.25967/650388

ground systems. This initiative brings together industry experts, regulatory bodies, and academic researchers. Notable outputs include the "Artificial Intelligence in Aeronautical Systems: Statement of Concerns" (ER-022 / AIR6988) [8] and the forthcoming "Process Standard for Development and Certification Approval of Aeronautical Products Implementing AI" (ED-324 / ARP6983) [9]. The ED-324 / ARP6983 will offer extensive guidance on integrating machine learning into aeronautical systems, covering data management, model design, validation, verification, implementation, and certification activities. The initial focus of ED-324 / ARP6983 will be on offline supervised learning. It is designed to interface with other standards like ED-79B/ARP 4754B [10] for airborne systems development, or ED-12C/DO-178C [11] for software items and ED-80/DO-254 [12] for complex hardware items EASA also follows the progress of other working groups on AI, in particular ISO/IEC SC42 [13] and CEN CENELEC JTC21 [14].

2.2. Military

On the other hand, in Europe, military and state aircraft are excluded from the scope of the EASA legislation. Consequently, military airworthiness is only regulated at national level. The need to improve European defence capabilities influenced the decision of the European Defence Agency (EDA) to create the Military Airworthiness Authorities (MAWA) forum to develop, adopt and implement harmonised European Military Airworthiness Requirements (EMARs). At the same time, the EDA developed the European Military Airworthiness Certification Criteria (EMACC) to harmonise military airworthiness standards across Europe. It serves as a comprehensive guide for Military Airworthiness Authorities to establish tailored certification bases for new military aircraft and components by mapping existing military and civil standards (like MIL and EASA CS/DS) to a common set of high-level, qualitative criteria.

In order to discuss the main aspects for developing trustworthy AI Systems for European defence, the Trustworthiness for AI in Defence (TAID) Working Group, led by EDA and composed by AI experts from the defence industry, academia, and national Ministries of Defence has published a white paper [15]. The purpose of this document is to identify related regulations and standards, highlight the key requirements for developing trusted AI systems for defence use, to recommend research activities, and to create the foundation for future standards for AI defence. This effort is performed in the context of the EDA Action Plan on AI in Defence and tries to address the topics of trusted AI and verification, validation and certification requirements analysis. In parallel, the EU has provided funding and co-funding for various projects, including initiatives such as the European Initiative for Collaborative Air Combat Standardisation (EICACS) [16] and AI for Defence consortium (AI4DEF) [17].

When using EMACC, EASA AI specifications may not be enough. Future military standards may have to be developed to address AI-specific requirements, emphasising mission criticality, combat robustness and ethical considerations (see FIG 1).

In parallel, the North Atlantic Alliance (NATO) released NATO's revised AI Strategy 2024 [18] which builds upon the foundation laid in 2021 and emphasizes the responsible development and use of AI technologies in defence and security. As such, it reaffirms the six Principles of Responsible Use (PRU) for AI: Lawfulness, Responsibility and Accountability, Explainability and Traceability, Reliability, Governability, and Bias Mitigation. These principles, aligned with NATO's values, norms and international law, guide the Alliance's approach to AI adoption, ensuring ethical compliance and addressing potential risks.

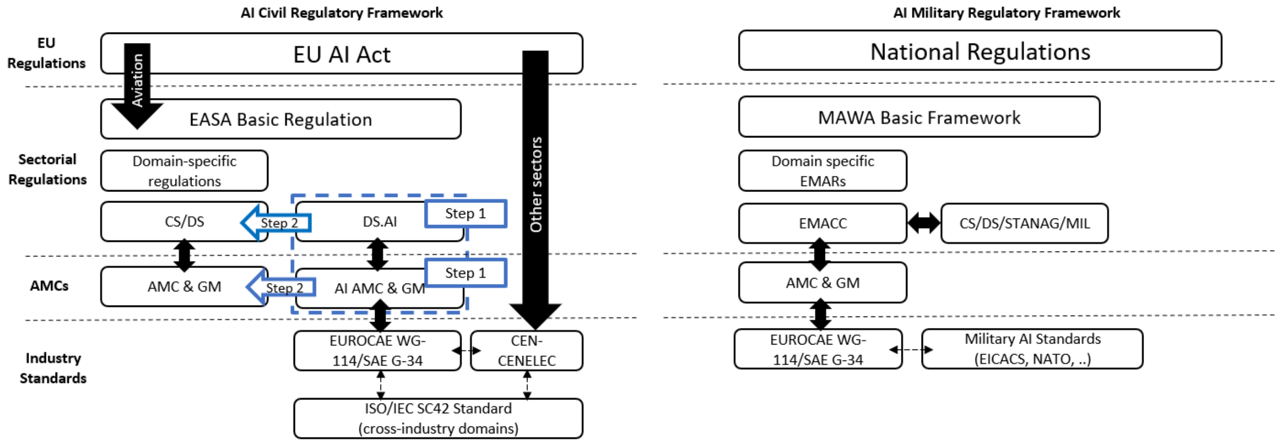

FIG 1. AI civil and military regulatory framework.

3. LEARNING ASSURANCE AND W-SHAPED PROCESS

The W-shaped process was introduced by EASA and DAEDALEAN during the first CODANN report in 2020 [19]. EASA reused it in their Concept paper: “Guidance for Level 1 & 2 machine learning applications” issue 2 [5] (see FIG 2). This W-shaped describes the development process of machine learning based systems. The main change compared to the traditional V-shaped development process is the introduction of the learning phase to the development assurance process, which is required so that ML models capture the intended function. The W-shaped process can be seen as a concatenation of two V-shaped processes, as illustrated in FIG 3. The first V, also called Machine Learning Development Lifecycle (MLDL), includes the learning phase of the model, whereas the second V (MLC implementation V) focuses on the implementation. Both V-shaped processes are described in the subsections 3.2 and 3.3.

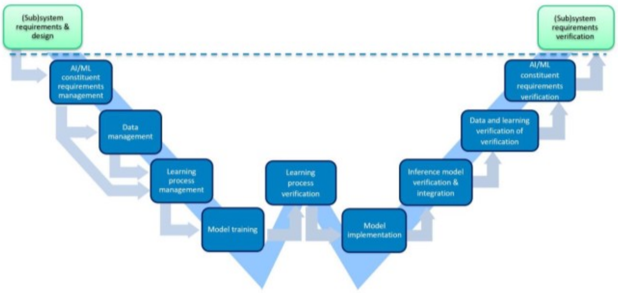
(Sub)system requirements & design
AI/ML constituent requirements management
Data management
Learning process management
Model training
Learning process verification
Model implementation
Inference model verification & integration
Data and learning verification of verification
AI/ML constituent requirements verification
(Sub)system requirements verification

FIG 2. W-shaped process. Source: [5, p.50].

The EUROCAE WG-114/SAE G-34 joint committee has worked on this W-shaped process and documented it in the ARP6983/ED-324 [9], describing each subprocess. They are grouped in 3 main sections:

- Section 4, covering system and safety considerations (see section 3.1);
- Section 5, covering the ML design phase (see section 3.2);
- Section 6, covering the implementation (see section 3.3).

In order to describe the transition from system to item, the machine learning constituent (MLC) was introduced: it corresponds to the lowest level of the system decomposition which is developed using ML, including the necessary data processing. The MLC operational design domain (ODD) is also an artifact used in the document to describe the refinement of the operating environment at the level of the MLC.

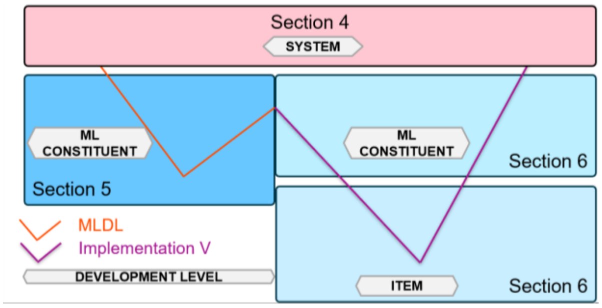

FIG 3. W-shaped process from EUROCAE WG-114/SAE G-34. Source: [9, p.24].

3.1. System and Safety Considerations

This section includes four main objectives:

- MLC development is defined within the system processes;
- Existing safety processes are used to assess MLCs;
- Operating environment is defined to a level of detail necessary to support development of MLC operational design domain (ODD);
- System-level performance requirements are established.

3.2. Machine Learning Development Lifecycle (MLDL)

This section includes five processes of the MLDL (ML Model design V):

- MLC requirements and logical architecture process;
- Data management process;
- ML model design process;
- ML validation process;
- ML verification process.

3.3. MLC Implementation, Integration and Verification

This section includes three processes (implementation V):

- The MLC Implementation Architecture Development Process;
- The MLC Items Implementation and Verification Process;
- The MLC Integration and Verification Process.

4. MILITARY USE CASES

In this section, the Enhanced Automatic Aerial Refuelling (UC-A) and Automatic Target Recognition (ATR) (UC-B) use cases will be used to assess the applicability of the W-shaped process on the development of military AI-based systems. In the sequence, these use cases will be described and the considerations and concerns related to each process of the first V (MLDL) of the W-shaped process are discussed, in particular the requirements identification and the architecture definition.

4.1. UC-A - Enhanced Automatic Aerial Refuelling

4.1.1. Description

Aerial refuelling (AR) is an essential military capability in which a tanker aircraft transfers aviation fuel to a receiver aircraft using a boom or hose-and-drogue system while flying in close formation. Given the complexity of this task, Airbus Defence & Space has developed and certified the Automatic Aerial Refuelling Boom System (ARBS). This system is designed to perform automatic operations, thereby reducing the air refuelling operator's (ARO) workload, enhancing safety, and optimizing the rate of air-to-air refuelling transfer.

The Automatic ARBS can be enhanced in terms of versatility, particularly in complex backgrounds, and further automated through the use of machine learning. For example, a convolutional neural network (CNN) trained with supervised learning can be employed to support the detection of the receiver aircraft in images captured by the AR cameras. This specific use case will be analysed in this paper.

4.1.2. System and Safety Considerations

In this process, following the ED-79B/ARP4754B [10], the system function related to providing the receiver aircraft's position is assigned to an MLC called ML Detection and Estimation (MLDE), and this MLC is defined within the system architecture. Additionally, a set of system functional, performance, availability, and interface requirements are allocated to it. Some of these requirements are shown in TAB 1.

Next, the operating environment of the MLDE MLC is identified, represented as operational requirements (see TAB 1), based on the Concept of Operations (ConOps), and system and environmental characteristics.

Based on the assigned function and the Concept of Operations, a safety assessment is performed following the ED-135/ARP4761A [20]. This resulted in the allocation of Development Assurance Level (DAL) D to the MLDE MLC and the establishment of safety requirements (S-SA-01 and S-SA-02 shown in TAB 1).

TAB 1. Example of Allocated Requirements.

ID	Requirement description
S-F-01	The ML Detection and Estimation shall detect and identify the receiver aircraft approaching the tanker using the coordinates of the image received from the AR cameras.
S-F-02	The ML Detection and Estimation shall calculate the coordinates of the upper-left and bottom-right vertices of the bounding box that encloses the receiver aircraft in the image, considering a coordinate system where: - The upper-left corner of the image is considered as (0, 0) (the origin point); - The upper-right corner of the image is considered as (W, 0); - The bottom-left corner of the image is considered as (0, H); - The bottom-right corner of the image is considered as (W, H), where W represents image width and H represents image height.
S-OP-01	The ML Detection and Estimation shall provide detection and state estimation for receiver aircraft of type F-15 and F-16.
S-OP-02	The ML Detection and Estimation shall operate in conditions where the minimum receiver's relative roll angle is *min_rel_roll* degrees and the maximum relative roll angle is *max_rel_roll*.
S-OP-03	The ML Detection and Estimation shall operate in conditions of clear and cloudy sky.
S-OP-04	The ML Detection and Estimation shall operate in conditions in which the tanker flies over sea and desert.
S-SA-01	The safety repercussions of the ML Detection and Estimation shall not be worse than MINOR.
S-SA-02	The ML Detection and Estimation shall be developed in accordance with Development Assurance Level (DAL) D requirements.
S-PE-01	The ML Detection and Estimation shall achieve a mean Average Precision (MAP) of at least *min_mAP* averaged over a range of Intersection over Union (IoU) thresholds from 0.5 to 0.95, with a step size of 0.05, on the designated test dataset.

4.1.3. MLDL

4.1.3.1. MLC Requirements and Logical Architecture Process

In the first MLDL process, those allocated system requirements shown in TAB 1 are refined with consideration given to specific aspects of the ML solution. This refinement includes detailing operational requirements, which compose the operational design domain (ODD), to facilitate data collection, verification, and monitoring. For example, the requirement S-OP-03 (TAB 1) can be refined as MLC-OP-01 (see TAB 2). In addition, robustness requirements are defined to ensure the system's reliability and performance under various conditions as presented in TAB 2.

TAB 2. Example of MLDE MLC Requirements.

ID	Requirement description
MLC-OP-01	The ML Detection and Estimation shall operate in the four different lighting conditions: cloudy morning, sunny morning, cloudy afternoon, and sunny afternoon.
MLC-OP-02	The ML Detection and Estimation shall detect the receiver aircraft when it is placed in any region of the image.
MLC-RO-01	The ML Detection and Estimation shall be robust to situations in which the boom entirely or partially obscures the receiver aircraft.
MLC-RO-02	The ML Detection and Estimation shall be robust to the presence of different cloud types in the background.

In the sequence, the logical architecture of the MLDE MLC is defined. This architecture is composed of three elements: pre-processing, CNN detector, and post-processing. These elements are illustrated in FIG 4 and described in TAB 3.

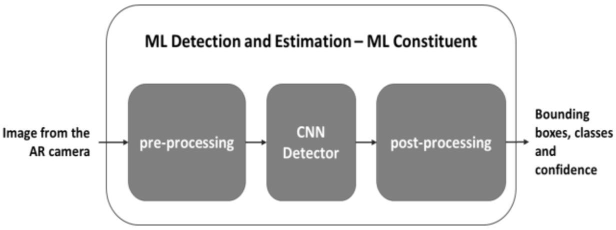
ML Detection and Estimation – ML Constituent
Image from the AR camera
pre-processing
CNN Detector
post-processing
Bounding boxes, classes and confidence

FIG 4. MLDE MLC Logical Architecture.

TAB 3. Elements of the MLDE MLC Logical Architecture.

MLC element	Input	Description	Output
Pre-processing	Image from AR Camera.	It resizes and normalises the original image.	Resized and normalised image.
CNN detector	Resized and normalised image.	It processes the images and provides bounding boxes indicating the presence of the receivers on the resized image in the local coordinate system.	Bounding boxes indicating the position of the receivers on the resized image, receiver aircraft types (classes), and related confidence score.
Post-processing	Bounding boxes indicating the position of the receivers on the resized image.	It transforms the bounding boxes from the coordination system of the resized image to the one of the original image.	Bounding boxes indicating the position of the receivers on the original image, receiver aircraft types (classes) and related confidence score.

After the MLC requirements and the logical architecture are defined, the next step of the W-shaped process is the Data Management.

4.1.3.2. Data Management Process

The data consists of images captured by AR cameras in a specific format, which need to be resized (down-scaled) and normalised during the operations according to the input requirements of the convolutional neural network (CNN). One type of output of the CNN is the bounding boxes enclosing the receiver aircraft on the resized image, which must be up-scaled to match the size of the original image. The other types of outputs are the aircraft types found in each bounding box and the respective confidence score.

The existing real images are insufficient to cover the operational design domain (ODD), so synthetic data needs to be generated to create complete datasets to be used in training and verification (e.g., training, validation, test-generalisation, test-robustness, and test-uncertainty quantification datasets). Various types of test datasets were considered to account for the different data characteristics required to evaluate the ML model's generalization capability, robustness, and uncertainty quantification.

For this use case, the following data sources are identified:

- Flight test videos: Videos captured with AR cameras during aerial refuelling flight tests.
- Synthetic data generator: Simulation environment with 3D flight models of the tanker and receiver aircraft.

To guide the data source selection, data generation/collection, preparation, and allocation into datasets, data quality requirements (DQRs) are defined based on the MLDE MLC requirements, as exemplified in TAB 4. For instance, based on the requirement MLC-OP-01 about lighting conditions (TAB 2), the requirements DQ-06, DQ-07, and DQ-08 were created.

TAB 4. Example of DQRs for the MLDE MLC.

ID	Requirement description
DQ-01	The images from the training, validation, and all test datasets shall be in the format 8 bits BGR and size *Width* x *Height* x *Colour*.
DQ-02	Each image shall be allocated exclusively to one of the datasets.
DQ-03	The images from the training, validation, and all test datasets shall be captured by either a real or a synthetic aerial refuelling camera.
DQ-04	The synthetic data source shall incorporate the F-15, F16 and boom visual and physics models with realistic 3D meshes and textures.
DQ-05	The training, validation, test-generalisation, and test-uncertainty datasets shall contain images with F-15s, F-16s, and without any aircraft, considering the following percentages: - 30% of the images in the dataset shall contain only F-15s. - 30% of the images in the dataset shall contain only F-16s. - 30% of the images in the dataset shall contain F-15s and F-16s. - 10% of the images in the dataset shall not contain any receiver aircraft.
DQ-06	The training, validation, test-generalisation, and test-uncertainty datasets shall contain images with the four specified lighting conditions: cloudy morning, sunny morning, cloudy afternoon, and sunny afternoon. Each lighting condition shall represent 25% of the dataset samples in each dataset.
DQ-07	The images shall include information about the lighting condition in their metadata.
DQ-08	The images shall be classified regarding their lighting conditions based on the mean brightness, computed as the mean of the V (Value) component for the image coded in the Hue, Saturation, Value (HSV) colour space.

After the allocation of the datasets, the training and validation datasets together with a data processing description are provided to the ML Model Design process.

4.1.3.3. ML Model Design Process

The process begins with defining specific ML model requirements based on the MLC requirements. These requirements include specifying the input format as resized and normalized images from AR cameras, defining the output format as bounding box coordinates enclosing the receiver aircraft, and selecting an appropriate loss function to optimize during training like cross-entropy loss (classification loss), and Complete Intersection over Union (CIoU) loss and Distribution Focal Loss (DFL) (localization losses).

Following the requirements definition, a CNN model architecture is selected and trained using prepared training and validation datasets. The model's parameters are adjusted through backpropagation to minimize the defined loss functions and improve performance.

After training, a detailed model description document is created. This document includes a clear description of the data that the model takes as input and the output it produces. It also includes an explanation of the CNN architecture, outlining its hyperparameters (e.g., layers and their types, learning rate, batch size), and the parameters of each layer (e.g., number of filters, filter size, stride). Additionally, the document offers an analytical/algorithmic description, detailing the algorithms and mathematical operations used within the model.

The developed artifacts from Data Management and ML Model Design processes will then be either validated or verified.

4.1.3.4. ML Validation Process

The MLDE MLC requirements, DQRs, ML model requirements, MLC logical architecture, and ML Model logical architecture are validated through reviews and data analysis. The validation ensures that these requirements are accurate, consistent, verifiable, and conform to standards. Additionally, it is validated that:

- MLC requirements comply with system requirements;
- DQR, DPRs and ML Model requirements comply with MLC requirements;
- DPRs comply with DQRs;
- MLC logical architecture is compatible with MLC requirements;
- ML Model logical architecture is compatible with ML Model requirements;
- MLC and ML Model logical architectures are consistent and comply with standards.

4.1.3.5. ML Verification Process

In the ML Verification Process, both the data and the ML model are verified to ensure adherence to MLC, data quality, and ML model requirements. Data verification involves assessing various dimensions of data quality, including completeness, representativeness, accuracy, timeliness, and uniqueness. For the ML model, the evaluation focuses on several critical aspects. To ensure generalisation, the model's performance is evaluated on unseen data using a dedicated test-generalisation dataset. Robustness is assessed by testing the model's resilience to noise and adversarial examples (e.g., clouds and occlusion by the boom). Uncertainty quantification methods are employed to estimate the model's predictive uncertainty. Additionally, explainability methods are used to identify potential biases.

In addition to the prepared test datasets, scenario-based testing aligned with the ODD is also employed. For this purpose, the synthetic data generator described in Section 4.1.3.2 was utilized.

4.1.4. Military Considerations

For this use case development, the W-shaped process and its objectives could be applied. However, certain military characteristics, such as mission secrecy, can pose challenges and necessitate increased customer involvement. For instance, the customer might not be able to share all operational details, leading to an incomplete or incorrect definition of the operating environment. This, in turn, can result in an incomplete or incorrect operational design domain (ODD) and robustness requirements, which are crucial for creating appropriate datasets. Consequently, the model's ability to generalise and exhibit robust behaviour may be compromised, as it was not trained and verified with the correct datasets or against the appropriate requirements.

Moreover, secrecy can also lead to a lack of real data, making it difficult to accurately characterize the operating environment and ensure the representativeness of datasets. This scarcity of real data can further hinder the model's ability to generalise effectively in real-world scenarios.

Therefore, a potential solution could be to establish a collaborative framework that includes the customer in the training activities, as well as to make use of synthetic data generation.

4.2. UC-B - Automatic Target Recognition (ATR)

4.2.1. Description

Performing intelligence, surveillance, and reconnaissance (ISR) operations with Unmanned Aircraft Systems (UASs) is crucial for modern militaries, providing real-time battlefield tracking, and situational awareness. The complex and adversarial environments in which these operations are conducted necessitate advanced technologies to ensure high performance and robustness. ML offers a promising solution to enhance ISR capabilities, especially in the area of ATR. Therefore, this use case proposes the improvement of ATR with the use of an onboard ML model that is capable of accurately identifying, classifying, and tracking military-relevant targets in real-time. The ML model should be capable of detecting and classifying various objects such as:

- Vehicles (e.g., military trucks, ships, tanks);
- Infrastructure (e.g., military bases, command centres, airfields);
- Weapons systems (e.g., missile launchers, artillery).

4.2.2. System and Safety Considerations

The system to be analysed is the ISR system of a UAS which incorporates a ML model, specifically a CNN trained with supervised learning. This ML model will be part of a MLC called machine learning-based automatic target recognition (MLATR), which will also contain a pre-processing element.

The MLATR MLC will support the system function of identifying and classifying the following targets: vehicles, infrastructure, and weapons systems. This information will then be provided to a human operator who will be responsible for the decision-making process.

To achieve this, system functional, operational, robustness, safety, interface, and performance requirements are allocated to this MLC, as exemplified in TAB 5, according to the ED-79B/ARP4754B [10].

Based on the assigned function and the Concept of Operations, a safety assessment is performed following the ED-135/ARP4761A [20]. This results in the allocation of DAL E to the MLDE MLC and the establishment of a safety requirement (see TAB 5).

It is important to note that while the MLATR MLC operation does not inherently imply safety effects, the authors recommend using the learning assurance framework, which is provided by the W-Shaped Process, to develop this type of military system and ensure best practices. This approach enhances system availability and effectiveness, which are critical for mission success.

TAB 5. Example of Allocated Requirements.

ID	Requirement description
S-F-01	The MLATR MLC shall detect and classify the following military elements: vehicles, infrastructure, and weapons systems.
S-OP-01	The MLATR MLC shall provide detection and classification for vehicles of type: military trucks, ships, and tanks.
S-OP-02	The MLATR MLC shall provide detection and classification for infrastructure of type: military bases, communication towers, and airfields.
S-OP-03	The MLATR MLC shall provide detection and classification for weapon systems of type: missile launchers and artillery.
S-OP-04	The MLATR MLC shall operate in desert regions, mountainous and high-altitude terrain, urban environment, coastal environment, and forested regions.
S-OP-05	The MLATR MLC shall operate in the four seasons of the year.
S-OP-06	The MLATR MLC shall operate during day and night.
S-RO-01	The MLATR MLC shall be robust against decoys which mimic real targets.
S-SA-01	The MLATR MLC shall not cause safety repercussions.
S-IN-01	The MLATR MLC shall receive input data coming from Infrared (IR) sensors.
S-IN-02	The MLATR MLC shall output bounding boxes enclosing the targets, along with the identified class and confidence score, for each detected target.
S-PE-01	The MLATR MLC shall achieve a mean Average Precision (mAP) higher than *min_mAP* at IoU higher than 0.5 across all target classes, as tested

After the definition of the system and safety considerations, the MLDL can be started.

4.2.3. MLDL

4.2.3.1. MLC Requirements and Logical Architecture Process

In the first MLDL process, the allocated system requirements are refined with consideration given to specific characteristics of the ML solution, as shown in TAB 6.

In the sequence, the MLATR MLC logical architecture, as represented in FIG 5, is defined. This architecture consists of two main constituent elements:

- Pre-processing: This element is responsible for resizing the images received from an Infrared (IR) sensor.
- ML Detector: This element is in charge of detecting the presence and classifying military targets in the image using a ML model.

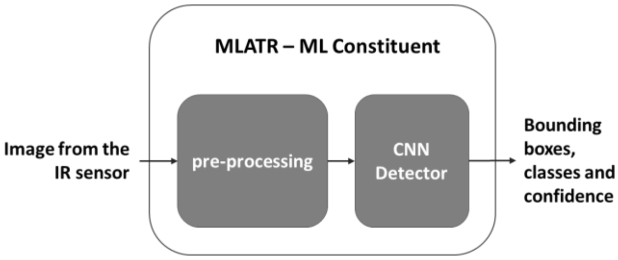

FIG 5. MLATR MLC Logical Architecture.

TAB 6. Example of MLATR MLC Requirements.

ID	Requirement description
MLC-OP-01	The MLATR MLC shall operate in the 3 different time of day conditions: day, night, and transition.
MLC-OP-02	The MLATR MLC shall operate in the five environmental conditions: snow, forest, sea, mountains/rocks, and urban.

The MLC requirements and logical architecture are then provided to the ML Data Management and Model Design processes.

4.2.3.2. Data Management Process

For this use case, images from an IR sensor are used as input to the ML model. To resize the image according to the size required by the ML model, a pre-processing step is established.

The output of the ML model is defined as bounding boxes with the respective target class and confidence score.

With the type of input and output defined, data quality requirements (DQRs) are identified, as exemplified in TAB 7. Addressing the DQRs, real images are collected and synthetic images are generated. These images are then prepared and allocated to the datasets, including training, validation, test-generalisation, test-robustness, and test-uncertainty quantification datasets.

TAB 7. Example of DQRs for the MLATR MLC.

ID	Requirement description
DQ-01	The images from the training, validation, and all test datasets shall be in the *desired_image_format* and size *Width* x *Height*.
DQ-02	Each image shall be allocated exclusively to one of the datasets.
DQ-03	The images from the training, validation, and all test datasets shall be captured by real or synthetic IR sensors.
DQ-04	The synthetic data source shall incorporate three target types: vehicles, infrastructures and weapon systems.
DQ-05	The training, validation, test-generalisation, and test-uncertainty datasets shall contain images with the three target types (vehicles, infrastructures and weapon systems), and without any target, considering the following percentages: - 12% of the images in the dataset shall contain only target vehicles. - 12% of the images in the dataset shall contain only target infrastructure. - 12% of the images in the dataset shall contain only target weapon systems. - 12% of the images in the dataset shall contain target vehicles and infrastructure. - 12% of the images in the dataset shall contain target vehicles and weapon systems. - 12% of the images in the dataset shall contain target infrastructure and weapon systems. - 12% of the images in the dataset shall contain target vehicles, infrastructure, and weapon systems. - 16% of the images in the dataset shall not contain any target.
DQ-06	The training, validation, test-generalisation, and test-uncertainty datasets shall contain images with the three specified time of the day conditions: day, night, and transition. Each time of the day condition shall represent 33,3% of the dataset samples in each dataset.
DQ-07	The images shall include information about the time of the day condition in their metadata.
DQ-08	The training, validation, test-generalisation, and test-uncertainty datasets shall contain images with the five environmental conditions: snow, forest, sea, mountains/rocks, and urban. Each environmental condition shall represent 20% of the dataset samples in each dataset.
DQ-09	The images shall include information about the environmental condition in their metadata.
DQ-10	10% of the images from the training, validation, test-generalisation, and test-uncertainty datasets shall contain decoys.
DQ-11	The images shall include information about the presence of decoys in their metadata.
DQ-12	40% of the images from the test-robustness dataset shall contain decoys.

4.2.3.3. ML Model Design Process

In this process, the ML model requirements are defined, the model is built, trained, optimised, and documented in the ML model description document. The ML model requirements, as mentioned in the previous use case, describe the format of input, the selected algorithms, the loss function of the training, training stopping criteria, the metrics to evaluate the training, and the output format.

The resizing of the image will be selected according to model performance requirements (e.g., latency) and implementation constraints.

The selected ML model is a CNN that is trained to fulfil the performance and robustness requirements identified in TAB 5.

Once the MLC, ML Data and ML Model requirements, and logical architectures are defined, the ML Validation Process starts.

4.2.3.4. ML Validation Process

The MLATR MLC requirements, DQRs, ML model requirements, MLC logical architecture, and ML model architecture are validated through reviews and data analysis.

The next step is the verification of the datasets and ML Model against these requirements.

4.2.3.5. ML Verification Process

The data and ML model verification process addresses the same objectives as those outlined for the Enhanced Automatic Aerial Refuelling (UC-A) use case, as mentioned in Section 4.1.3.5. Additionally, the process includes specific robustness verification steps using adversarial examples to ensure the ML model's resilience against potential threats and countermeasures.

4.2.4. Military Considerations

For this use case development, the W-shaped process and its objectives could also be applied. However, mission secrecy may pose challenges in defining the operating environment, ODD, and robustness requirements. These challenges can compromise the model's ability to generalise, exhibit robust behaviour, and be verified against the correct requirements, as discussed in Section 4.1.4.

Moreover, the complex and adversarial environment in which this ML model is expected to operate imposes extra difficulties in defining the ODD. The enemy may employ various tactics to confuse the ML model, such as using decoys to mimic the appearance of real targets or using infrared jammers to disrupt the model's inputs.

To mitigate these issues, the following measures could be implemented:

- Adversarial training and verification techniques to improve the ML model's robustness.
- Continuous system updates to adapt the ML model to new threats and countermeasures. This could mean updating the model offline and undergoing a new certification and/or qualification process, as well as online learning opportunities (not currently addressed by ARP6983/ED-324 [9]). To facilitate faster updates, adjustments to the framework outlined in ED-324/ARP6983 may be

required, as the frequency of updates may exceed that of typical civil applications.

- Use of synthetic data to generate a variety of scenarios and adversarial examples, which are not covered by the real data, to improve the model's ability to generalise and its robustness against adversarial tactics.

5. CONCLUSION

The analysis of the two use cases confirm that the learning assurance concept proposed in the ED-324/ARP6983 [9] is highly relevant for the development of military applications. Examples of requirements were provided for different steps of the W-shaped process.

The specificities of military applications were highlighted. In particular, the mission secrecy resulting in a degraded definition of the operating environment, leading to an incorrect ODD, and to robustness and generalisation issues. Moreover, military operations imply very fast changing operating conditions, which again leads to a difficult definition of the ODD and an outdated ML model.

To address these specificities, several recommendations were provided. These include proposing a collaborative framework to involve the customer (who is the end-user) in the training of the model, as well as focusing on mitigation measures such as: performing adversarial training; applying robustness verification techniques; using synthetic data; and continuously training the ML model to provide quick system updates. Some of these measures might require adaptations to the framework outlined in ED-324/ARP6983 [9].

6. ACKNOWLEDGMENT

The authors would like to thank the EUROCAE WG-114/G-34, as well as the following colleagues for their review of this paper: Adrián Jiménez González, Antonio Monzon Diaz, Avelino Martin Adalid, Christophe Gabreau, Nicolas Valot, and Christoph Stahl.

7. DISCLAIMER

The use cases presented in this paper are original and have been developed ad hoc for illustration purpose.

This document describes a proposed standard that is currently a work in progress and subject to ongoing development, review, and revision. The concepts, data, and conclusions presented herein are preliminary and may evolve as technical discussions continue. Due to the complex and technical nature of standards development, the only official and enforceable version of the standard will be the final, formally published document. This paper is provided solely for informational purposes and should not be relied upon as a definitive statement of the final standard or its requirements.

8. SYMBOLS

Symbol	Description
desired_image_format	Desired format of the image.
min_mAP	Minimum mean Average Precision threshold.
min_rel_roll	Minimum relative roll angle between receiver and tanker aircraft measured in degrees.
max_rel_roll	Maximum relative roll angle between receiver and tanker aircraft measured in degrees.

9. ACRONYMS

Acronym	Definition
AI	Artificial Intelligence
AMC	Acceptable Means of Compliance
ANS	Air Navigation Services
AR	Aerial Refuelling
ARBS	Automatic Target Recognition
ARO	Air-to-air Refuelling Operation
ATM	Air Traffic Management
ATR	Automatic Target Recognition
CEW	Cognitive Electronic Warfare
CNN	Convolutional Neural Network
ConOps	Concept of Operations
CS	Certification Specification
DAL	Design Assurance Level
DFL	Distribution Focal Loss
DQR	Data Quality Requirements
DS	Detailed Specifications
EASA	European Union Aviation Safety Agency
EDA	European Defence Agency
EMACC	European Military Airworthiness Certification Criteria
EMAR	European Military Airworthiness Requirements
EU	European Union
F	Functional
GM	Guidance Material
HSV	Hue, Saturation, Value
IN	Interface
IoU	Intersection over Union
IR	Infrared
MAWA	Military Airworthiness Authorities forum
MCRI	Military Certification Review Items
MIL	Military Standard
ML	Machine Learning
MLATR	ML-based Automatic Target Recognition
MLC	ML Constituent
MLDE	ML Detection and Estimation
MLDL	Machine Learning Development Lifecycle
NATO	North Atlantic Treaty Organization
ODD	Operational Design Domain
OP	Operational
PE	Performance
RMT	Rule Making Task
RO	Robustness
SA	Safety Assessment
SAE	Society of Automotive Engineers
SARP	Standards and Recommended Practices
SC	Special Condition
STANAG	NATO Standardization Agreement
PRU	Principle of Responsible Use (NATO)
UAS	Unmanned Aerial Systems
UC	Use Case

10. REFERENCES

[1] Monzon Diaz, A., Capdevila Llompart, C. M., “Particularities of Certifying Artificial Intelligence in Military Aviation,” 2024.

[2] European Union Aviation Safety Agency (EASA), "Artificial Intelligence Roadmap 2.0 – A human-centric approach to AI in aviation," 2023.

[3] European Commission, “Regulation (EU) 2024/1689 - AI Act,” 2024.

[4] European Union Aviation Safety Agency (EASA), "EASA Concept Paper: guidance for Level 1 machine learning applications," 2021.

[5] European Union Aviation Safety Agency (EASA), "EASA Concept Paper: guidance for Level 1 & 2 machine learning applications," 2024.

[6] European Commission, “Regulation (EU) 2018/1139 - EASA Basic Regulation,“ 2018.

[7] European Union Aviation Safety Agency (EASA), "Terms of Reference for Rulemaking Task - ToR RMT 0742," Issue 1, 2024.

[8] EUROCAE and SAE, "Artificial Intelligence in Aeronautical Systems: Statement of Concerns," 2021.

[9] EUROCAE and SAE, "[DRAFT 7] ED-324/ARP6983 - Process Standard for Development and Certification Approval of Aeronautical Products Implementing AI," EUROCAE and SAE," 2024.

[10] EUROCAE and SAE, “ED-79B/ ARP4754B Guidelines for Development of Civil Aircraft and Systems,” 2023.

[11] EUROCAE and RTCA, “ED-12C/DO-178C Software Considerations in Airborne Systems and Equipment Certification,” 2012.

[12] EUROCAE and RTCA, “ED-80/ DO-254 Design Assurance Guidance for Airborne Electronic Hardware,” 2000.

[13] International Organization for Standardization, “ISO/IEC JTC 1/SC 42 - Artificial Intelligence,” 2018.

[14] CEN CENLEC. “Joint Technical Committee 21 (JTC 21),” 2021. Available in: https://jtc21.eu/

[15] European Defence Agency (EDA), “Trustworthiness for AI in Defence – White Paper,” 2025.

[16] European Defence Fund. “EICACS: European Initiative for Collaborative Air Combat Standardisation,” 2021. Available in: https://defence-industry-space.ec.europa.eu/system/files/2023-01/Factsheet_EDF21_EICACS.pdf

[17] AI4DEF. “About us,” 2024. Available in: https://ai4def.com/about-us/

[18] NATO (2024). “Summary of NATO’s revised Artificial Intelligence (AI) strategy,” Available in: https://www.nato.int/cps/en/natohq/official_texts_227237.htm

[19] EASA and Daedalean AG. “Concepts of Design Assurance for Neural Networks,” 2020. Available in: https://www.easa.europa.eu/sites/default/files/dfu/EASA- DDLN- Concepts- of- DesignAssurance-for-Neural-Networks-CoDANN.pdf.

[20] EUROCAE and SAE, “ED-135/ARP4761A Guidelines for Conducting the Safety Assessment Process on Civil Aircraft, Systems, and Equipment,” 2023.

[21] Chen, Hai-Wen, and Ravi Kapadia. "Improving automatic target recognition (atr) performance with electro optics (eo) and infrared (ir) sensor fusion," 2022 IEEE Aerospace Conference (AERO). 2022.