TOWARD A REGULATORY FRAMEWORK TO CERTIFY ARTIFICIAL INTELLIGENCE APPLICATIONS IN AVIATION

Dr. Antonio Monzon Diaz, CSEP
Expert in Aerospace Development and Certification
Airbus Defence and Space
Rechliner Str., 85077 Manching
+49 (0) 151-16724926
antonio.monzon-diaz@airbus.com

Abstract

In response to the increasing demand for the integration of Artificial Intelligence (AI) technology in aviation, the regulatory agencies and international standardization bodies are working together to pave the way for the establishment of a regulatory framework to certify AI applications in aviation. Specifically, the European Union Aviation Safety Agency (EASA) and European Organization for Civil Aviation Equipment (EUROCAE), together with Society of Automotive Engineers (SAE), are developing regulatory materials and standards within the European environment. This paper aims to provide an overview of these ongoing initiatives and identify the primary challenges that must be addressed to ensure the effective adoption of AI technology by the aviation industry.

1. INTRODUCTION TO ARTIFICIAL INTELLIGENCE (AI) AND MACHINE LEARNING (ML)

The concept of Artificial Intelligence (AI) has been proposed since the early stages of the computer science, however its practical development did not happen immediately. In the last decade, this technology has significantly progressed due to the increased capacity to manage massively data together with the exponential increase in computing power, and supported by the development of powerful algorithms. AI is a broad field of computer science which intends to mimic human intelligence by computer systems. Machine Learning (ML) is a subset of AI which deals with the capability of a computer to “learn” from existing data without someone having to explicitly program its behavior [11]. Learning means using the data to find values for the parameters of an algorithm to try to reach the expected behavior.

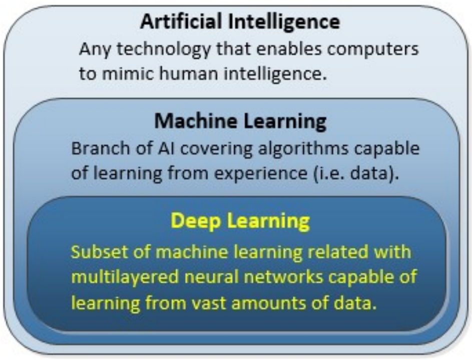

Figure 1. Machine Learning and Deep Learning as part of Artificial Intelligence

The most common algorithms used in machine learning are Artificial Neural Networks (ANN) and decision trees. ML is typically divided into three subcategories based on the type of data available: Supervised Learning, Unsupervised Learning and Reinforcement Learning.

1.1. Use of ML in aviation

ML algorithms are particularly efficient for pattern recognition (e.g. natural language recognition, automated computer vision, etc.). In particular computer vision is one of the areas where ML technology has shown great potential for aviation application. Some examples of how ML-based computer vision can be used in aviation are:

- Object detection: ML algorithms can be used to detect objects in images captured by cameras installed on aircraft. This can include runway lights, other aircraft, drones or obstacles in general. By using computer vision techniques, ML models can accurately identify these objects, providing pilots with real-time information to help avoid collisions and other hazards.

- Autonomous on-ground operation during taxi: Computer vision can also be used to help aircraft move autonomously on-ground. ML models can be trained to recognize landmarks, such as runways, taxiways, and other features of the airport environment, allowing aircraft to move accurately and safely without human intervention.

- Predictive maintenance: ML can also be used to predict when maintenance will be required on aircraft parts or mechanical systems. By analyzing data from sensors and other sources, ML models can identify crack patterns that indicate when a structural component is likely to fail, allowing maintenance personnel to replace it before it causes a problem.

1.2. Challenges of ML towards certification

The development of ML applications represents a change of paradigm with respect to the traditional software or hardware as it inverts the traditional sequence of generating the expected behavior by implementing fixed rules (manual process). In ML paradigm the expected behavior is in fact an input used to train the algorithms so that they can automatically generalize and obtain the rules.

In addition to this change of paradigm ML applications show a number of characteristics which make them especially difficult to certify for its use in aviation.

- Lack of transparency: ML applications are often considered “black boxes”, meaning that it can be difficult to understand how they make decisions. This lack of transparency can make it challenging to verify that the system is operating as intended and to identify and address any potential safety issues derived from unintended behavior.

- Lack of predictability: In essence ML algorithms are by nature probabilistic (i.e. their results have to be interpreted in probabilistic terms), however they are implemented as any other algorithm and as such they provide the same outputs for the same inputs, and consequently they can be considered formally deterministic. On the other hand, when a new input is given to the ML model, the output may depend on the correlation between that input and the data set used during the training phase. For this reason, this output can be considered a priori as unpredictable. Additionally, ML models may provide significantly different outputs for small variations of inputs (especially for classification applications).

- Complexity: ML models can be extremely complex, with many layers of algorithms and parameters. This complexity makes it difficult to fully test and validate the system, as there may be many different scenarios and corner cases that need to be considered.

1.3. Need of new regulatory framework for AI

Given the challenges and novel development paradigm detailed in the preceding section, it is clear that there is a pressing need to establish a new regulatory framework that can effectively address the unique characteristics of certifying Machine Learning applications.

In particular EASA and EUROCAE/SAE are developing specific regulatory material and standards to support the development and certification of AI-based applications in aviation. The rest of this paper aims to present the state of the science in this field by providing a concise summary of the various concept papers published by EASA, in addition to elucidating the strategy of EUROCAE/SAE toward establishing a novel industrial standard for this domain.

2. EASA CONCEPT PAPERS ON AI

EASA has published the concept papers “EASA AI Roadmap 1.0” [1], "First usable guidance for Level 1 machine learning applications” [3] and "First usable guidance for Level 1 & 2 machine learning applications” [4]. Related to this, EASA has also issued the Special Condition "SC-AI-01 Trustworthiness of Machine Learning based Systems" [9], which prescribes the use of the "First usable guidance for Level 1 machine learning applications”.

The purpose of this section is to summarize this EASA material and identify the main underlying concepts proposed by EASA to address the specificities of certifying ML-based applications.

2.1. EASA AI Roadmap 1.0

The EASA AI Roadmap 1.0 [1] provides a comprehensive and structured approach for integrating AI-based systems into the aviation industry while ensuring the highest levels of safety, security, privacy, and ethical responsibility. The roadmap is based on four key pillars: safety, security, privacy, and ethical considerations. To ensure safety, the roadmap emphasizes the need for AI applications to be certified and validated, and for proper risk management procedures to be implemented. To address security concerns, the roadmap advocates for the development of secure and resilient systems, as well as the establishment of clear guidelines for cybersecurity. Privacy is addressed by promoting the responsible use of data and ensuring compliance with data protection regulations. In general, ethical considerations are emphasized through the development of ethical standards and guidelines for the development and use of AI applications.

The EASA AI Roadmap 1.0 introduces the notion of level of autonomy which refers to the ability of AI systems to operate independently and make decisions without direct human intervention. The roadmap recognizes that autonomy is a critical component of many AI applications in aviation, particularly in unmanned aerial systems (UAS). EASA proposes three levels of autonomy for AI systems in the aviation industry, which are defined as follows:

1. Level 1 AI/ML: Assisted Operations (assistance to human) - The system is capable of assisting the human operator in making decisions, but the final decision remains with the human.

2. Level 2 AI/ML: Semi-Autonomous Operations (human/machine collaboration) - The system is capable of making some decisions independently, but a human operator is still required to oversee and intervene if necessary.

3. Level 3 AI/ML: Fully Autonomous Operations (autonomous machine) - The system is capable of making all decisions independently and does not require a human operator, however the human is expected to be in the loop at the design and oversight phases.

In the current regulatory framework, the associated risk-based approach for systems, equipment and implementation items are mainly driven by a requirement-based development assurance methodology during the development of their constituents. EASA recognizes that elements based on learning processes cannot be addressed with this development assurance scheme due to their intrinsic nature, and for this reason the EASA AI Roadmap 1.0 also introduces the concept of learning assurance. Learning assurance is based on the idea that ML systems are per nature capable to learn and evolve their behavior based on the data and inputs they receive. To ensure that these systems continue to operate safely and effectively, it is necessary to have ongoing monitoring and validation processes in place to detect and correct any errors or issues that may arise. This way EASA promotes the creation of novel means of compliance with objective is to gain confidence at an appropriate level that an AI application supports the intended functionality. The definition of learning assurance provided in [3] is as follows:

> "*All of those planned and systematic actions used to substantiate, at an adequate level of confidence, that errors in a data-driven learning process have been identified and corrected such that the system satisfies the applicable requirements at a specified level of performance, and provides sufficient generalisation and robustness guarantees*".

Additionally the EASA AI Roadmap 1.0 introduces the concept of AI explainability which refers to the "*capability to provide the human with understandable, reliable, and relevant information with the appropriate level of details and with appropriate timing on how an AI/ML application produces its result*". This is particularly important in the aviation industry, where the use of AI systems can have significant implications for safety. The need for AI explainability arises from the fact that many AI systems use complex algorithms and decision-making processes that may be difficult for humans to understand. This can create challenges in terms of accountability, transparency, and trust, as users may be reluctant to rely on AI systems that they cannot fully understand.

2.2. EASA AI Roadmap 2.0

The EASA AI Roadmap 2.0 [2] keeps the overall structure and strategy already proposed in the previous version and introduces the following main changes:

- Extension of the scope of AI techniques (see Fig. 2): in addition to the ML and DL techniques covered in the previous version, logic- and knowledge-based (e.g. expert systems), statistical techniques (e.g. Bayesian estimation) and hybrid techniques (i.e. combinations of different approaches) have been added.

- Reference to European Union AI Act: this version establishes a clear link with the proposed legislation EU AI Act which is expected to become an EU law after approval of the EU member states.

- Refinement of AI autonomy levels: the level 3 is now splitted into 2 subcategories 3A "The AI-based system performs decisions and actions that are overridable by the human" and 3B "The AI-based system performs non-overridable decisions and actions".

- Anticipated rulemaking concept for AI: EASA presents a two-steps approach for the future establishment of the regulatory framework on AI, being the first step the establishment of a dedicated Part-AI (at the same level as the existing Part 21). The ongoing industrial standard ARP6983 is also mentioned.

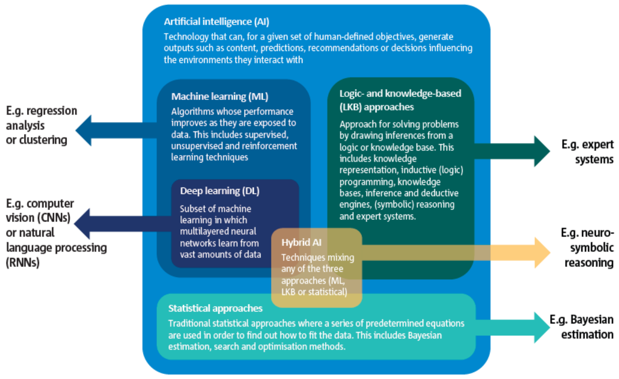

Figure 2. Scope of technology covered by AI Roadmap 2.0 [2]

2.3. EASA Concept Paper "First usable guidance for Level 1 machine learning applications"

The EASA Concept Paper "First usable guidance for Level 1 machine learning applications" [3] provides a first set of objectives for the development, validation, and certification of Level 1 machine learning applications in aviation. The paper does not constitute a definitive or detailed guidance material; instead it serves as basis for the future development of formal regulatory material. In particular, it supports supervised learning strategy. Future evolutions would enlarge the ML techniques scope.

EASA proposes the notion of AI trustworthiness which intends to cope with the general European Commission ethical guidelines aviation context by the implementation of different higher level objectives organized in four building blocks (see Fig. 3).

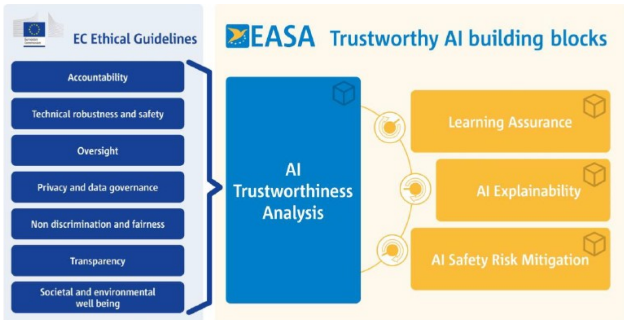

Figure 3. EASA AI trustworthiness roadmap building blocks [1]

These building blocks are mapped into the following sections, which propose dedicated objectives per topic:

2.3.1. Trustworthiness analysis

The trustworthiness analysis aims to ensure that the application is safe, secure and ethics-compliant. These three aspects are considered by EASA important prerequisites in the development of any system that uses or embeds AI/ML, and consequently are integral processes towards the approval of AI-based applications.

The trustworthiness analysis starts with a characterization of the AI application, which involves understanding the application's purpose, features, operational limitations and assumptions. This initial step is crucial to identify potential risks and challenges that may arise during the development and deployment of the application.

The guidance emphasizes that the trustworthiness analysis encompasses both safety and security assessments. The safety assessment evaluates the application's ability to perform its intended function safely and effectively, while the security assessment evaluates the application's vulnerability to cybersecurity threats. The most relevant objectives in these areas are:

> *"**Objective 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.*
>
> ***Objective IS-01**: For each AI-based (sub)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."*

Additionally the guidance emphasizes that the trustworthiness analysis should also include an ethics-based assessment, which intends to evaluate the ethical implications of the AI application. This assessment aims to ensure that the AI application does not violate ethical principles or cause damages to individuals or the society as a whole. The most relevant objective in this area is:

> *"**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."*

2.3.2. AI assurance

The AI assurance proposes a system-centric approach to developing AI-based systems. This approach is complemented by an end-user centric approach that considers human factors related to AI. The objectives for the AI-based system are defined, considering the unique aspects of machine learning techniques. The concept of learning assurance is introduced, with a focus on data management and learning processes. The lack of transparency of ML models is addressed through explainability objectives (see example below).

> *"**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 (including assumptions made on the level of training, qualification and skills)."*

Data-recording objectives are established to support continuous safety assessment, performance monitoring, and investigation in case of an incident or accident during the operations of the AI-based system.

2.3.2.1. Learning assurance

The learning assurance concept aims to ensure that an AI-based system can perform its intended function at a satisfactory level of performance (among other contributions). This concept also ensures that the trained models used to produce the system are reliable and can handle new data beyond what was used during the training process. Essentially, the learning assurance concept aims to guarantee that the AI-based system will work well in the real world, and will be able to adapt to new situations not specifically used during the training of the model.

The steps of the learning assurance process are inspired in the classical V-shape development assurance with some adaptation to ML concepts to allow structuring the learning assurance guidance. EASA proposes a W-shaped process framework to represent the ML lifecycle (see Fig. 4).

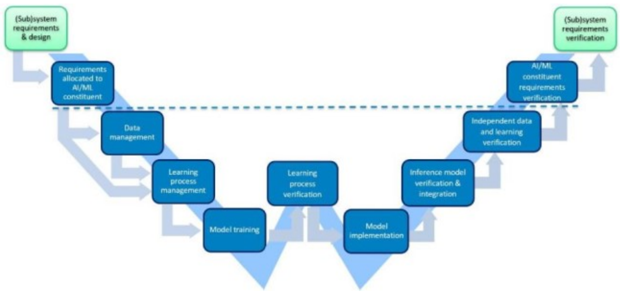

Figure 4. Learning assurance W-shaped process [3]

EASA proposes dedicated objectives per lifecycle activity, for example:

> "***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.*"

2.3.2.2. Development & post-ops AI explainability

Post-operational phases refer to the stages of an aircraft's life cycle after it has been put into service, e.g. maintenance, repair or overhaul. During the post-operational phases, EASA requires that the aircraft and its components be maintained and repaired in accordance with approved procedures and standards to ensure continued airworthiness and safety. EASA also requires that operators of aircraft maintain accurate records of all maintenance and repairs performed on the aircraft, as well as any modifications or upgrades made to the aircraft's systems or components. This information is used to track the aircraft's history and ensure that it remains in compliance with all applicable regulations and safety standards.

In the context of AI constituents, this section aims to establish a link between AI assurance and development/post-ops explainability by providing a framework for achieving transparency on the ML model. Although the learning assurance process may not offer complete transparency on the inner workings of complex models such as NNs, associated explainability methods can support the objectives of learning assurance and operational explainability.

The most relevant objective in this section is:

> "***Objective EXP-03***: *The applicant should identify and document the methods at AI/ML item and/or output level satisfying the specified AI explainability needs.*"

2.3.3. AI safety risk mitigation

The concept of AI Safety Risk Mitigation (SRM) acknowledges that complete opening of the 'AI black box' may not always be possible or practical, which could result in residual risks. SRM is intended to minimize the likelihood of unintended or unexplainable outputs by implementing mitigations such as real-time monitoring of the AI-based system with recovery through traditional backup systems. The use of AI in aviation is a relatively new domain, and until sufficient field experience is gained, appropriate safety precautions should be taken to minimize risks to population and critical infrastructure.

The main objective proposed by EASA in this section is:

> *"**Objective 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 whether an additional dedicated layer of protection, called hereafter safety risk mitigation (SRM), would be necessary to mitigate the residual risks to an acceptable level."*

2.3.4. Human Factors for AI

The EASA Concept Paper "First usable guidance for Level 1 & 2 machine learning applications" [3] replaces the building block AI Explainability by the new block "Human Factors for AI", however the AI Assurance concept remains within this building block.

With this change EASA acknowledges the importance of Human Factors in the overall AI explainability concept, recognizing it as the main driver to achieve an effective understanding of the behavior of a ML-based application.

Additionally EASA introduces dedicated Human Factor objectives (with respect to the previous version [3]):

> *"**Objective HF-01**: The applicant should design the AI-based system with the ability to build its own individual situational awareness.*
>
> ***Objective 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.*"

2.4. SC-AI-01 Trustworthiness of Machine Learning based Systems

A Special Condition (SC) is an additional airworthiness standard prescribed by EASA when the relevant airworthiness regulations do not properly cover the corresponding safety requirements due to novel or unusual design features.

In the context of ML-based application, EASA has issued the Special Condition "SC-AI-01 Trustworthiness of Machine Learning based Systems" [8] in order to assist manufacturers, developers, and operators in identifying and mitigating the risks associated with ML-based systems. In order to fulfill this objective the SC-AI-01 prescribes the use of the "First usable guidance for Level 1 machine learning applications" to supplement the Certification Specifications CS 23.2500 & 23.2510 for applications embedding AI/ML technology, as these paragraphs do not provide adequate CS for ML-based systems, in particular the aspects related with AI explainability and Ethics-based assessment.

The SC-AI-01 can be considered as an interim guidance reference until dedicated AI regulatory material is formally produced.

3. EUROCAE/SAE ARP6983

EUROCAE and SAE are collaborating through the joint international committee WG-114 / G-34 to develop the industrial standard ARP6983 "Process Standard for Development and Certification of Aeronautical Safety-Related Products Implementing Artificial Intelligence" [8]. The purpose of ARP6983 is to provide a standard process for the development and certification of aeronautical safety-related products that implement AI technology. It intends to support the development of safe and reliable AI-based products, while also providing a consistent process framework for certification authorities and industry stakeholders.

The EUROCAE/SAE ARP6983 is an international standard, currently under development, which proposes guidelines for the development and certification of Aviation Systems implementing AI/ML capabilities. It provides practices in different areas with special focus on the particularities of ML applications, e.g. ML data management or ML training management.

The ARP6983 shows a similar structure as the ARP4754 [7] or the ED-12C/RTCA DO-178C [5]: it proposes a novel Machine Learning Development Lifecycle (MLDL) in terms of processes, which is broken down into lower level objectives, inputs, activities and outputs, and additionally proposes different objectives linked to the activities to be fulfilled depending on the DAL allocated to the ML constituent. This development lifecycle is based on the W-shaped process framework proposed by EASA (see Fig. 4).

The ML development lifecycle starts with the ML Constituent Operational Design Domain (MLCODD) process. The MLCODD process takes system/subsystem requirements and plans/standards applicable to MLDL as inputs, and produces a description of the MLCODD as output. The next step is the ML Constituent Requirements Process, which develops both ML Constituent requirements and ML Constituent derived requirements. This process takes inputs such as system/subsystem requirements, architecture, MLCODD process outputs, and plans/standards applicable to MLDL.

The next step is the Data Management process in Machine Learning (ML), which main objective is to provide high-quality training, validation, and test datasets for use in ML model training, design enhancement, and system integration to meet requirements and functional/operational needs. This process also includes delivering a data processing description for use in subsequent software or hardware requirements processes.

Taking the datasets as inputs, the next step is the ML Model Design Process. This process aims to iteratively develop the ML Model architecture and the ML Model requirements to be used to build, train, and optimize the ML Model.

Once the ML Model is produced, the next steps are the ML Validation and Verification Processes. The main goal of the validation is to determine that the requirements are correct and complete, and the aim of ML Verification is to ensure that both the ML Data and the ML Model adhere to the validated ML Constituent requirements. Verification may involve various activities such as reviews, analyses or tests. Once the ML Model Description (MLMD) is validated and verified, it is considered ready for implementation.

The ML constituent implementation is intended to be performed following the existing processes for the respective software and hardware items which represent the physical architecture of the ML constituent. At this stage the MLDL interfaces with the respective SW and HW development lifecycles as described in ED-12C/RTCA DO-178C [5] and ED-80/RTCA DO-254 [6] respectively.

Once the corresponding SW and HW items are developed, the process continues with the subsequent integration activities. The next step would then be the ML Model integration and finally the ML constituent integration. The final output of the MLDL would be the ML constituent integrated within its sub-system to become part of the system (product) including the ML constituent.

The standard proposes as well an AI-related terminology, describing all the concepts in this field which are relevant for the purposes of the standard.

The main novelties proposed by the ARP6983 are related with the introduction of new processes in the following areas: ML Data Management (including data gathering and preparation), ML Model Design (including ML training), ML Validation (dealing with the assurance of ML requirements completeness and correctness), ML Verification (dealing with the assurance that the ML Data and ML Model fulfills with the ML Requirements) and ML Implementation.

The ARP6983 can be considered as an intermediate engineering layer between the system lifecycle process (ARP4754) and the implementation lifecycle processes (e.g. DO-178C) dedicated to ML based system/subsystem. This new standard is expected to coexist with all the standards already in place (see Fig. 5), and to complement the ARP4754 for subsystems embedding ML technology.

4. FUTURE AVIATION REGULATORY FRAMEWORK FOR AI

EASA has still to decide how to incorporate the position papers in place as formal regulatory material. Additionally the industrial standard ARP6983 is still under development, and its first issue is not expected until 2024. For these reasons, the topic of fitting the new regulatory material and standard is still to be defined. However, it is expected that the European regulatory agency will follow a similar approach as for currently enforced standards and regulations.

Typically, Certification Specifications (CSs), along with Acceptable Means of Compliance (AMCs) and Guidance Material (GMs), ensure compliance with European civil aviation legislation (particularly EC 1139/2018 and EC 748/2012), which in turn comply with ICAO Standards and Recommended Practices (SARPs).

The European Commission's Artificial Intelligence Act (AI Act) is a proposed regulation aimed at governing the development and use of artificial intelligence (AI) for general purposes within the European Union (EU). The act outlines a comprehensive set of rules and requirements for AI. The AI Act aims to ensure that AI systems are designed and used in a way that is safe, transparent, and respects fundamental rights. The regulation includes provisions related to transparency and explainability, data protection, human oversight, and accountability. The proposed regulation is intended to create a framework for the development of trustworthy AI within the EU, while also promoting innovation and competitiveness. The AI Act is currently being reviewed by the European Parliament and the Council of the EU, and is expected to be adopted in the near future. This legislation constitutes the overarching legal framework that will be developed through the future Certification Specification to regulate the development and certification of AI-based systems in aviation.

The Special Condition SC-AI-01 represents a supplement to the existing certification specification CS 23 “Certification specifications for normal, utility, aerobatic, and commuter category airplanes”, and as such it would be at the same level as the rest of EASA regulatory material (see Fig. 5).

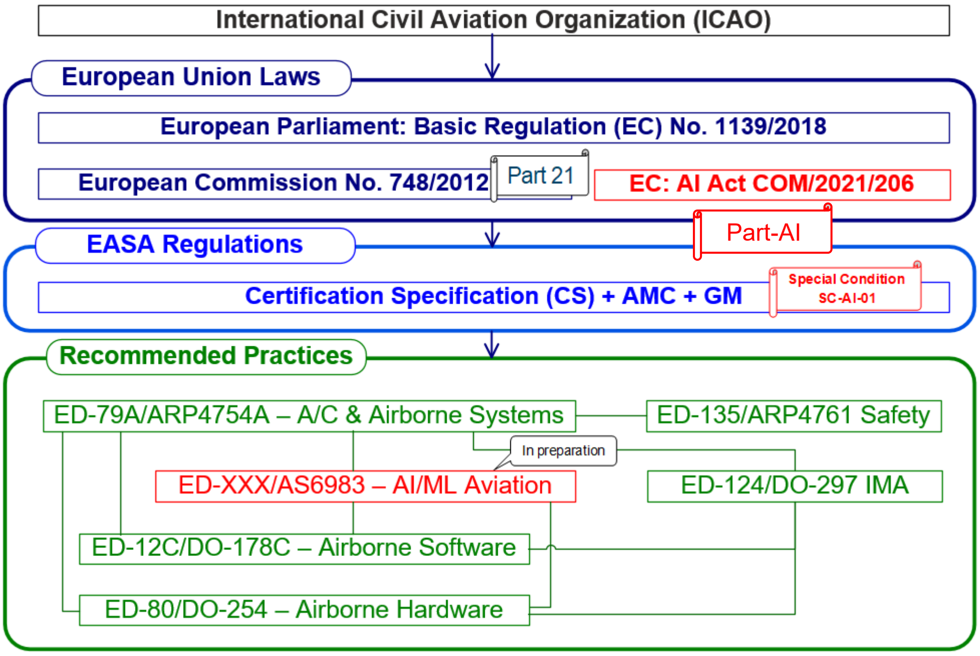

Figure 5. Aviation Regulatory Framework for AI applications

The new standard ARP6983 would be integrated with the rest of industrial standards. It can be expected that the ARP6983 (system/sub-system level) will be positioned between the ARP4754 and equipment-level standards (DO-178C and DO-254). In this way, the new AI standard is intended to coexist with current standards without interfering with traditional developments that do not include AI technology.

EASA has proposed the anticipated rulemaking concept for AI applications in aviation as part of the EASA AI Roadmap 2.0 [2] (see Fig. 6).

As part of this anticipated rulemaking concept, EASA has proposed a two-steps approach for the deployment of the new regulatory material:

- Step 1: Development of a transversal Part-AI containing the provisions anticipated in the EASA Concept paper guidance: requirements for authorities (Part-AI.AR), requirements for organizations (Part-AI.OR) and requirements on AI trustworthiness (Part-AI.TR).

- Step 2: Analysis of domain-specific requirements in order to assess the need to be complemented to provide an adequate regulatory basis for deploying the new Part-AI.

Taking this anticipated rulemaking concept into account, the future EASA AI regulatory framework supporting the certification of aeronautical systems embedding AI technology would basically consist in the new Part-AI (at the same level of existing Part 21), which would contain the requirements identified in the concept papers [3] and [4], organized in three major provisions: Part-AI.AR: requirements for authorities Part-AI.OR: requirements for organizations, and Part-AI.TR: requirements on AI trustworthiness.

Additionally EASA would propose in the future some dedicated acceptable means of compliance (AMC) and guidance material (GM) to supplement Part-AI requirements. At present, the only available guidance material for the certification of such systems is the Special Condition SC-AI-01.

Additionally, it can be expected that more detailed guidelines will emerge to aid the implementation of the new ARP6983 standard, similar to the RTCA DO-248 [10] which serves to assist in the deployment of the ED-12C/RTCA DO-178C standard. Another matter to consider in the future is the procedure for formally evaluating adherence to the process defined by the new ARP6983 standard. Software audits are currently used to demonstrate compliance with the objectives of the DO-178C standard (commonly referred to as "Stage of Involvement" or SOI). It would be reasonable to assume that similar audits may be established at AI constituent level to demonstrate that the development of a specific AI-based system has followed the process described in the new ARP6983 standard.

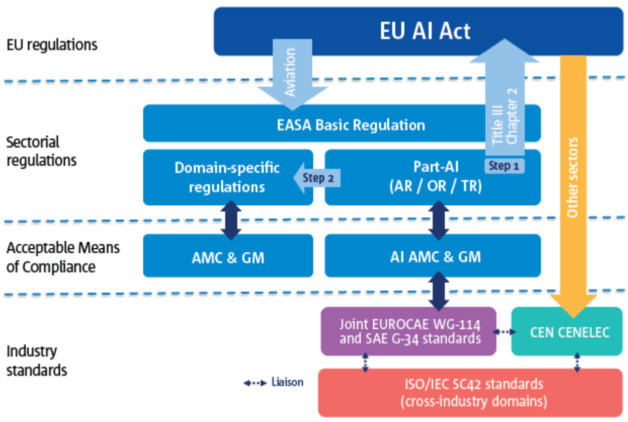

Figure 6. Anticipated regulatory structure for AI [2]

All of these considerations will need to be clarified and further developed in the future, and EASA will need to establish the corresponding Acceptable Means of Compliance which would serve to the applicants as guidance for the certification of AI-based applications in aviation.

5. CONCLUSIONS

This paper summarizes current regulatory material related with the certification of AI-based systems for use in aviation. It presents the main notions introduced by EASA's concept papers "EASA AI Roadmap 1.0", "EASA AI Roadmap 2.0", "First usable guidance for Level 1 machine learning applications", and "First usable guidance for Level 1 & 2 machine learning applications". Additionally, the paper briefly introduces the Special Condition SC-AI-01 Trustworthiness of Machine Learning-based Systems and the new international standard EUROCAE/SAE ARP6983.

The main challenges identified in EASA's concept papers regarding the certification of AI-based systems in aviation are the lack of transparency and predictability, and high complexity. These challenges are specifically addressed within the regulatory references, however the concrete techniques or methods to be used as novel demonstration means are still to be developed. Additionally these concept papers identify the higher level objectives to be fulfilled, which will be complemented by the lower level objectives identified in the new standard ARP6983.

The main contribution of this paper is to bring together the available material and to show how the new regulatory material for AI-based systems could fit into the existing aviation regulatory framework. In particular, it is indicated that in the future, a new regulatory framework is expected to be established by EASA (outlined in Figure 6) as reference for the industry when developing and certifying AI technology-based systems in aviation, which would be complemented by the process defined in the new EUROCAE/SAE ARP6983 standard. Until this dedicated regulatory material for AI is formally produced, the Special Condition SC-AI-01 can be used as an interim reference for guidance in the development and certification of AI-based systems.

Finally this paper anticipates two potential future lines of development in this field: additional detailed guidance material to assist in the application of the new ARP6983 standard and dedicated guidance for the assessment of the adherence of a system embedding AI-technology with this new standard.

Acknowledgments. I would like to send special thanks to Gary Brown, Christoph Gabreau, Avelino Martin, Janaina Ribas and Alexis de Cacqueray for the review of this paper and valuable feedback.

6. REFERENCES

[1] EASA AI Roadmap 1.0, 2020 (https://www.easa.europa.eu/en/newsroom-and-events/news/easa-artificial-intelligence-roadmap-10-published).

[2] EASA AI Roadmap 2.0, 2023 (https://www.easa.europa.eu/en/document-library/general-publications/easa-artificial-intelligence-roadmap-20).

[3] EASA First usable guidance for Level 1 machine learning applications, 2021 (https://www.easa.europa.eu/en/newsroom-and-events/news/easa-releases-its-concept-paper-first-usable-guidance-level-1-machine-0).

[4] EASA First usable guidance for Level 1 & 2 machine learning applications, 2023 (https://www.easa.europa.eu/en/downloads/137631/en).

[5] ED-12C/RTCA DO-178C: Software Considerations in Airborne Systems and Equipment Certification, RTCA Inc., 2011.

[6] ED-80/RTCA DO-254: Design Assurance Guidance for Airborne Electronic Hardware, RTCA Inc., 2000.

[7] SAE ARP4754 Rev. A: Certification Considerations for Highly Integrated or Complex Aircraft Systems, Society of Automotive Engineers Inc., 2010.

[8] SAE ARP6983: Process Standard for Development and Certification of Aeronautical Safety-Related Products Implementing Artificial Intelligence (in preparation, publication foreseen in 2024).

[9] Special Condition "SC-AI-01 Trustworthiness of Machine Learning based Systems", 2022 (https://www.easa.europa.eu/en/downloads/136369/en).

[10] RTCA DO-248: Supporting Information for DO-178C and DO-278, 2011.

[11] Russell, S. J., Norvig, P., Davis, E.: Artificial intelligence: a modern approach. 4th ed., Prentice Hall, 2020.

7. ACRONYMS AND TERMS

AI: Artificial Intelligence

AMC: Acceptable Means of Compliance

ANN: Artificial Neural Network

CRI: Certification Review Item

CS: Certification Specification

DAL: Design Assurance Level

EASA: European Union Aviation Safety Agency

EUROCAE: European Organization for Civil Aviation Equipment

GM: Guidance Material

ICAO: International Civil Aviation Organization

ML: Machine Learning

MLCODD: ML Constituent Operating Design Domain

MLDL: Machine Learning Development Lifecycle

MLMD: Machine Learning Model Description

NN: Neural Network

SAE: Society of Automotive Engineers

SARP: Standards and Recommended Practices

SC: Special Condition

SOI: Stage of Involvement

SRM: Safety Risk Mitigation

UAS: Unmanned Aerial Systems