# Initial Airworthiness Special Condition

## Trustworthiness of Machine Learning based Systems

Warning

This document contains links to pages containing EU law and/or to pages on the EASA website. You should not click on those links as those destination pages will not contain up to date and accurate descriptions of your rights and obligations. Please access up to date version of the applicable UK law on the CAA website here

| EASA European Union Aviation Safety Agency | **Special Condition** | Doc. No.: **SC-AI-01**<br>Issue : 1<br>Date : 19 April 2022<br>Proposed ☐ Final ☒<br>Deadline for comments: N/A |
|---|---|---|

| | | |
|---|---|---|
| **SUBJECT** | : | Trustworthiness of Machine Learning based Systems |
| **REQUIREMENTS incl. Amdt.** | : | CS 23.2500 & 23.2510 amdt 5 [1] |
| **ASSOCIATED IM/MoC** | : | Yes☐ / No ☒ |
| **ADVISORY MATERIAL** | : | EASA Concept Paper 'First Usable Guidance for Level 1 Machine Learning Applications' issue 01, Ref. [1], EUROCAE ED-79A/SAE ARP 4754A, AMC20-115D & EUROCAE ED-12C/RTCA DO-178C, AMC20-152A & ED-80/RTCA DO-254, AMC20-189 |

**INTRODUCTORY NOTE:**

The following Special Condition (SC) has been classified as important but since the technical content was already subjected to public consultation there is no need to subject this Special Condition to public consultation in accordance with EASA Management Board decision 12/2007 dated 11 September 2007, Article 3 (2.) which states:

"*2. Deviations from the applicable airworthiness codes, environmental protection certification specifications and/or acceptable means of compliance with Part 21, as well as important special conditions and equivalent safety findings, shall be submitted to the panel of experts and be subject to a public consultation of at least 3 weeks, except if they have been previously agreed and published in the Official Publication of the Agency. The final decision shall be published in the Official Publication of the Agency.*"

**IDENTIFICATION OF ISSUE:**

In the frame of recent certification projects, EASA has received proposal to embed a Neural Network trained through Deep Learning techniques as part of their system development.

This approach resonates with the work that EASA is conducting in the frame of the EASA AI Programme, aiming at implementing the actions of the EASA AI Roadmap 1.0. In this respect, a number of challenges have been identified when dealing with Artificial Intelligence / Machine Learning (AI/ML) techniques and a first set of usable guidance has been made available through the publication of the EASA Concept Paper 'First Usable Guidance for Level 1 Machine Learning Applications' issue 01 (further referred to as *Ref. [1]).*

Ref [1] : https://www.easa.europa.eu/downloads/134357/en

In the recent year the re-emergence of Deep Learning (DL) produced significant improvements for many problems in computer vision and natural language processing (NLP), enabling new use cases and accelerating AI adoption. Deep learning (DL) is a subset of Machine Learning (ML)[2] that emerged with the use of deeper

[1] The current certification specifications (here considering CS-23 amendment 5, paragraphs 23.2500 & 23.2510) do not provide adequate CS for the aspects pertaining to AI Explainability and to the Ethics-based Asessment, therefore it is considered necessary by EASA to issue a special condition due to the novelty of the topic.

[2] The ability of computer systems to improve their performance by exposure to data without the need to follow explicitly programmed instructions.

neural networks (NNs), leading to large improvements in performance. This is the reason why EASA AI Roadmap 1.0 and the Level 1 ML guidance are focusing on data-driven AI approaches.

Data-driven learning techniques are a major opportunity for the aviation industry but come also with a significant number of challenges with respect to the trustworthiness of ML and DL solutions. Here are some of the main challenges addressed through the first set of *Ref. [1]* objectives:

- Adapting assurance frameworks to cover learning processes and address development errors in AI/ML constituents;
- Creating a framework for data management, to address the correctness (bias mitigation) and completeness/representativeness of data sets used for the ML items training and their verification;
- Addressing model bias and variance trade-off in the various steps of ML processes;
- Elaborating pertinent guarantees on robustness and on absence of 'unintended function' in ML/DL applications;
- Coping with limits to human comprehension of the ML application behaviour, considering their
- stochastic origin and ML model complexity;
- Managing the mitigation of residual risk in 'AI black box' (the expression 'black box' is a typical criticism oriented at AI/ML techniques, as the complexity and nature of AI/ML models bring a level of opaqueness that make them look like unverifiable black boxes (unlike rule-based software); and
- Enabling trust by end users.

To address the challenges of data-driven learning approaches, EASA has developed an AI Trustworthiness framework organised around four major building-blocks (AI Trustworthiness Analysis, Learning Assurance, AI Explainability and AI Safety Risk Mitigation), according to which the EASA guidance is organised.

The current certification specifications (here considering CS-23 amendment 5, paragraphs 23.2500 & 23.2510) do not provide adequate CS for the aspects pertaining to AI Explainability and to the Ethics-based Asessment, therefore it is considered necessary by EASA to issue a special condition due to the novelty of the topic.

Note: The EASA Concept Paper in *Ref. [1]* had been opened for a consultation of 10 weeks on 21st April 2021 and its issue 01 published on 20th December 2021. This Concept Paper will be subject to further updates, as technology evolves and new elements of guidance becomes available. These updates will further be suject to public consultation.

Considering all the above, the following Special Condition is defined:

**Special Condition**

**Trustworthiness of Machine Learning based Systems**

Applicability: as defined in Chapter C of *Ref. [1]*:

The Applicant is requested to demonstrate compliance with each applicable objective from the Chapter C of *Ref. [1]* and to plan the AI-based system development accordingly.

An overview of the proportionality of the objectives depending on the classification and criticality of the AI-based system is indicated in the Chapter D of *Ref. [1]*. Chapter G of *Ref. [1]* provides a set of acronyms and definitions that are necessary to support the understanding of the guidance identified in its Chapter C.