Considering Security of Artificial Intelligence and Machine Learning Essential in Aviation

AIA Civil Aviation Cybersecurity Subcommittee

Stefan Schwindt – Chair (GE Aerospace)

Sean Sullivan – Vice Chair (The Boeing Company)

Working Group Membership:

Anup Raje	Honeywell
Wes Ryan	Northrop Grumman
Stefan Schwindt	GE Aerospace

Summary

With the 2023 release of several Artificial Intelligence and Machine Learning (AI/ML) applications, there has been a heightened interest across all industrial sectors in using these new tools. AI/ML has the potential to automate many tasks that previously were only achievable manually through humans as well as unlock new capabilities by utilizing many large and different data sources. However, the use of AI/ML introduces new threats as evident in existing applications. As a safety-critical industry with long lifecycles and a slow rate of change, AI/ML should only be deployed with appropriate safeguards. While Standards Development Organizations (SDOs) are developing standards for implementing AI/ML with protections against unintentional errors, the current work is not considering protections against intentional errors and attacks on AI/ML. It is considered imperative that the relevant working groups immediately include cybersecurity considerations so that the first issue of the AI/ML certification standards may allow approval of AI/ML in aviation for a safe and secure implementation and deployment.

Table of Contents

Table of Figures

1 Benefits of Artificial Intelligence and Machine Learning in Aviation

AI/ML offer many advances in technology. AI/ML has the ability to process extremely large data sets, revealing correlations and trends not easily found by experts. The possibility of machines understanding unstructured data, speech, and text enables machines to replicate functions currently necessitating humans while offering a higher level of automation, consistency, and precision. There are many potential applications for improving efficiency and safety of aviation. In their AI Roadmap 2.0, EASA explored conceivable use cases for AI/ML and proposed the steps necessary to achieve societal and regulatory acceptance. AI/ML can be used for more or fully autonomous flight operations, more efficient route planning, safe higher traffic density, automated remote runway inspections, and improved airplane and engine maintenance.

2 Certification and Approval Principles in Aviation

Safety remains the top priority in aviation. Aircraft parts and products are certified to satisfy this principle. The Design Approval Holders (DAH) must demonstrate an understanding of their designs and mastery of all tools, technologies, and processes used in realizing their product. The Type Certificate (TC) is only granted when regulatory authorities such as the Federal Aviation Administration (FAA) and European Union Aviation Safety Agency (EASA) have reviewed and accepted the evidence of design and ensured that it meets the set targets and objectives for ensuring safety. Like the authorization of a TC for the technical product, various forms of approvals are issued to organizations permitting them to operate after proving that the relevant activities and duties will be discharged safely. Examples of operational approvals include aircraft maintenance requiring licensing of staff, the organization demonstrating the right tools, processes, and resources to correctly and safely maintain aircraft, or the licensing of pilots based on proving competence on safely flying respective aircraft types.

2.1 Certification of Software for Safety

Software used for airborne applications and certain ground applications are certified using similar processes. Airborne software is certified using DO-178C/ED-12C by assigning Design Assurance Levels (DAL) according to the identified criticality of the software functions, where the highest DAL is appropriate for software with a potential catastrophic impact. DO-278A/ED-109A is used for ground software in the Air Traffic Management (ATM) and Communication, Navigation and Surveillance (CNS) space with assignment of Assurance Level (AL) 1 for software with a potential catastrophic impact on aircraft. In both standards, processes are established to ensure that software is developed and verified to minimize errors through human actions in design, implementation and verification.

Recognizing how automation can support software development, DO-330/ED-215 was established to qualify tools that are used to perform the processes of DO-178C/ED-12C or DO-278A/ED-109A. The qualification of tools provides assurance that the tools will perform processes to the same level as if they were performed by a human.

2.2 Certification for Cybersecurity

EASA Executive Decision ED 2020/006/R introduced formal rules for considering cybersecurity in the certification of aircraft parts and products. As part of this decision, CS 25.1319 was established and the industry standards DO-326A/ED-202A and DO-356A/ED-203A are recognized as Acceptable Means of Compliance in AMC 20-42. The FAA will be issuing equivalent formal rules in 2024. Until the formal rules are issued, PS-AIR 21.16-02 applies where special conditions are applied that match CS 25.1319.

The European Union has published a set of regulations known as Part IS, requiring information security be considered as part of organizational approvals. EU 2022/1645 requires design, production organizations, and airports protect against information security threats in their operations that may impact aviation safety. Similarly, EU 2023/102 requires the same of maintenance organizations, operators, and training organizations to secure their activities.

DO-326A/ED-202A and DO-356A/ED-203A jointly provide the means to certify aircraft parts and products. The documents describe the means to identify threats to an aircraft or system, while analogous to DAL and Security Assurance Levels (SAL) are defined. SALs provide the mechanism for assuring that the potential for vulnerabilities through intentional or unintentional means in architecture, development, design, implementation, and verification is minimized.

There is no current cybersecurity standards equivalent for ground software developed to DO-278A/ED-109A. Industry standards are being developed at a higher Information Technology (IT) and Operational Technology (OT) level to support Part IS, but these standards will not apply to the development of individual applications.

3 Cybersecurity Risks Related to AI/ML

As AI/ML is intended to be used to perform functions within aviation, there are risks to both safety and security. AI/ML is used in other industries and there has been ample demonstration of risks related to intentional and unintentional causes. It may be tempting to believe that a (near) defect-free implementation of the model is sufficient. However, with AI/ML, the source of safety risks is two-fold: the environment in which the AI/ML model is built as well as the data set used to train the AI/ML model. It is understood that intentionally tampering with the data set can lead to unnecessary bias in the model. Similarly, impacting the environment can cause the model to be skewed from its purpose. Therefore, the existing standards material for airborne products—DO-178C/ED-12C, DO-330/ED215, DO-326A/ED-202A and DO-356A/ED-203A—may support securing the AI/ML software application on the aircraft but will be insufficient for complete protection of the aircraft. Furthermore, there are no standards that support securing ground applications nor how to achieve an approval for operational functions being newly performed by AI/ML applications.

Risks shared with other types of applications—such as vulnerabilities in code processing external inputs—will largely be addressed by existing processes providing protections. AI/ML have unique risks that need to be considered and processes must be developed to provide protection against these risks. The threats for AI/ML are related to the inherent models in training and in the deployed environment. Current guidance does not provide any protection against data poisoning or bias introduction. Data poisoning refers to intentional skewing of data sets intended to train a model to impact its outputs. Bias introduction is a similar concept where unnecessary bias is introduced in a model's output rather than manipulating the complete model. The current industry guidance does not address environmental issues such as model logic threat or model bypassing. Model logic threat is the term for an attack aimed at extracting the deployed models for reverse engineering to influence future training. Model bypassing is a form of attack to create conditions to evade the invocation of a model in its entirety to alter system response.

Various studies have been published on vulnerabilities and attacks specific to AI/ML. Microsoft has published their classification of vulnerabilities related to AI/ML applications. OWASP has also published the following material for specific subclasses of AI/ML: a "*Top 10 for Large Language Model (LLM) applications*" and a "*Top 10 for Machine Learning.*" These publications may help the generation of guidance specific to aviation such that the processes for implementing and certifying AI/ML in aviation provides assurance that these applications are secure. The UK National Cyber Security Centre (NCSC) has also published material warning of the risks related to LLMs and proposes some steps that should be incorporated in design processes for mitigation.

RTCA has invited inputs to discuss options on adapting MOPS, MASPS, and other RTCA documents to best account for AI and other emerging technologies.

ENISA (European Union Agency for Cybersecurity) also expressed the challenges and gaps that exist today in handling the cyber threat landscape for AI and has established an ad hoc working group to do a study and to provide recommendations on addressing these challenges. For more details, refer to ENISA's paper on AI and cybersecurity published June 2023.

EASA published their envisaged roadmap for introducing AI/ML to aviation which describes multiple levels of AI/ML and some initial concept papers on guidance for implementing these levels. These levels identify to which degree AI/ML can be successively entrusted with aviation functions and what aspects need to be considered to approve and/or certify these applications. EASA specifically highlights that cybersecurity is one fundamental component of approving aviation AI/ML. The FAA has published their own draft roadmap for AI/ML in aviation. While the FAA's roadmap does not specifically address cybersecurity, it shares many aspects of EASA's roadmap. The FAA states that AI/ML may only be initially applied to low-risk applications. Use of AI/ML should follow the same principles as all other design aspects in aviation with safety as a prime. The FAA also suggests that design needs to consider failure of AI/ML and protect against this—such as against improper learning.

On October 30, 2023, President Biden issued Executive Order 13690 on ensuring the safe and secure development of AI. The Order directs National Science Foundation (NSF) and National Institute of Standards and Technology to conduct research and develop standards. The Sector Risk Management Agencies (SRMA) are directed to evaluate sectorial and cross-sector risks and consider means to mitigate. Because of this Executive Order, input material may be generated that is useful for the aviation sector although standards and regulations must be adapted appropriately.

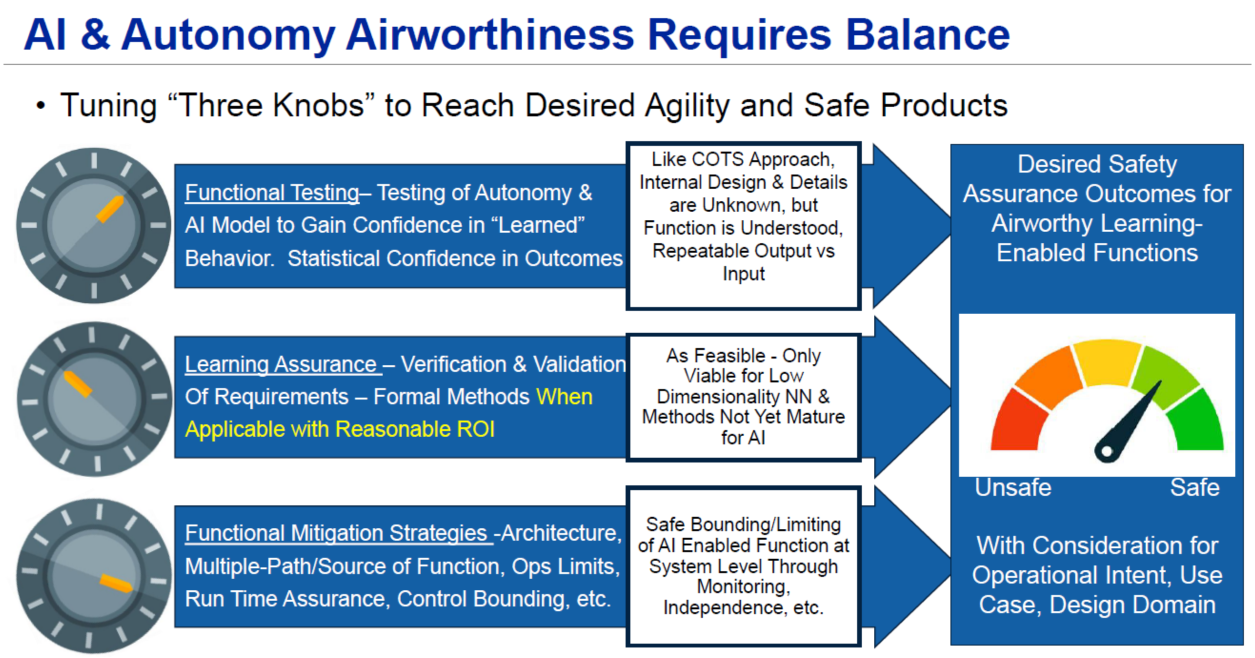
On November 27, 2023, 18 countries including all of the G7 group issued joint guidelines on securing AI systems and published them on the UK NCSC website. This is part of a concerted action to highlight the need for specifically securing AI systems for responsible deployment.

4 Ensuring Proportionality to Risks

As a general principle in aviation and standard practice in traditional safety, proportionality of measures to risks must maintained. This ensures that the available resources are allocated appropriately to such that more effort is expended to address greater concerns. This not only ensures that cost-benefit ratio of activities is met and that high severity risks receive attention, but it also ensures that human factors are addressed such that complacency does not arise. In traditional safety, the use of allocating and implementing Design and Development Levels (DAL) ensures proportionality where the higher DAL such as DAL A is used where Catastrophic Failure Conditions exist and DAL D for Minor Failure Conditions. The traditional security approach in aviation similarly utilizes Security Assurance Levels (SAL) with SAL 3 implemented to protect against more severe threat conditions.

In developing guidance for certifying AI/ML, similar levels of assurance should be introduced for aspects unique to AI/ML. Figure 1 shows three vectors along which assurance levels can be defined to achieve appropriate outcomes based on identified risks. While the diagram shows learning assurance, dimension includes verification and validation of requirements. This should also be understood to include the safeguarding of training data to protect against cybersecurity risks. Similarly, Figure 2 shows some of the technical and functional aspects that must be considered in proportionate controls of AI/ML. It is essential to identify the full risk of AI/ML applications, and as discussed in both EASA and FAA roadmaps, to ensure that a stepped approach within initially implementing low-risk solutions before gaining maturity to implement higher risk AI/ML. As part of the risk assessment for AI/ML, the human element must be considered as trustworthiness of AI/ML is important in both augmenting humans through supporting functions, as well as eventually replacing humans.

AI/ML may be implemented where there currently is no certified software, while assurance and trustworthiness is established through the training and licensing of human operators and maintainers. These aspects must be understood and considered in assuring safe, secure, and trustworthy AI/ML solutions.



AI & Autonomy Airworthiness Requires Balance

- Tuning "Three Knobs" to Reach Desired Agility and Safe Products

Functional Testing– Testing of Autonomy & AI Model to Gain Confidence in "Learned" Behavior. Statistical Confidence in Outcomes

Like COTS Approach, Internal Design & Details are Unknown, but Function is Understood, Repeatable Output vs Input

Learning Assurance – Verification & Validation Of Requirements – Formal Methods When Applicable with Reasonable ROI

As Feasible - Only Viable for Low Dimensionality NN & Methods Not Yet Mature for AI

Functional Mitigation Strategies -Architecture, Multiple-Path/Source of Function, Ops Limits, Run Time Assurance, Control Bounding, etc.

Safe Bounding/Limiting of AI Enabled Function at System Level Through Monitoring, Independence, etc.

Desired Safety Assurance Outcomes for Airworthy Learning-Enabled Functions

With Consideration for Operational Intent, Use Case, Design Domain

Figure 1 Assurance vectors for AI/ML [Wes Ryan, Northrop Grumman Corporation: AIAA SciTech 2024]

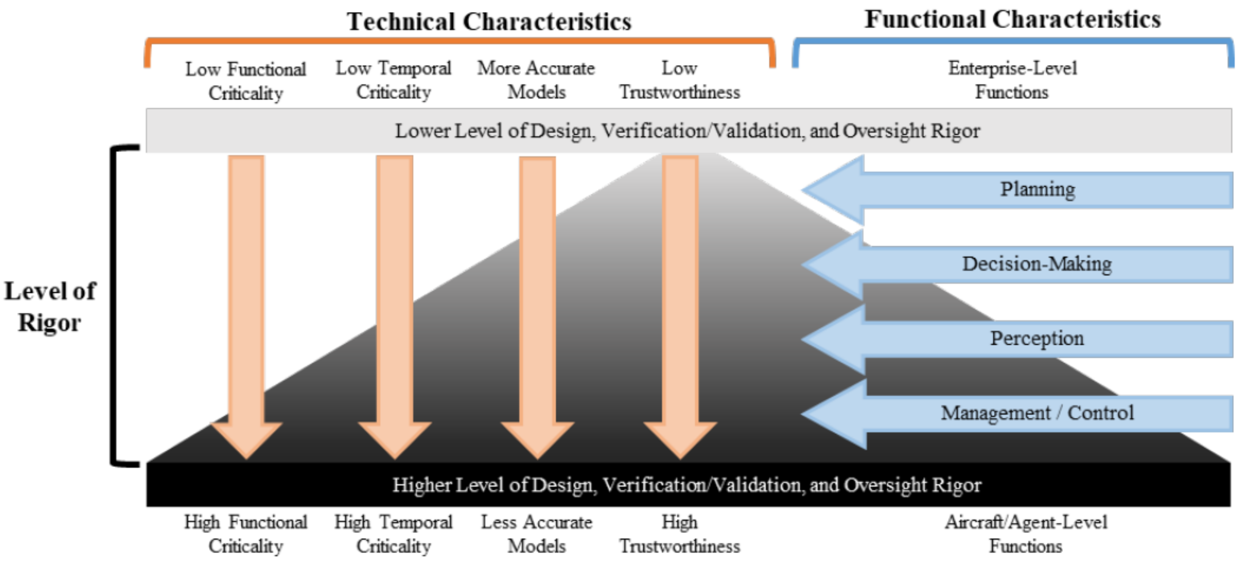

Technical Characteristics
Functional Characteristics
Low Functional Criticality
Low Temporal Criticality
More Accurate Models
Low Trustworthiness
Enterprise-Level Functions
Lower Level of Design, Verification/Validation, and Oversight Rigor
Level of Rigor
Planning
Decision-Making
Perception
Management / Control
Higher Level of Design, Verification/Validation, and Oversight Rigor
High Functional Criticality
High Temporal Criticality
Less Accurate Models
High Trustworthiness
Aircraft/Agent-Level Functions

Figure 2 Pyramid of rigor for AI/ML [INCS 2022]

5 Next Steps for AI/ML in Aviation

The AIA Civil Aviation Cybersecurity Subcommittee considers it a necessity for consideration of cybersecurity to allow the certification or approval of AI/ML applications. It is imperative that the SDOs SAE G-34 and EUROCAE WG-114 commence work on incorporating cybersecurity guidance into their standards material. Without such guidance, it may be impossible to demonstrate compliance with regulatory and societal requirements that applications are secure from vulnerabilities impacting safety. Without such evidence, certification or approval cannot be granted. Failure to include relevant guidance in the standards material jeopardizes the timely deployment of future applications by industry.

The AIA Civil Aviation Cybersecurity Subcommittee recommends that SAE G-34 and EUROCAE WG-114 review the threats and vulnerabilities associated with AI/ML development, deployment, and usage within aerospace applications and ensure that protections against these risks are included in the first issue of their standard proposed for certification. The guidance should consider how to protect models when trained or deployed on aircraft and on the ground, in addition to protecting the environment used for model training. Furthermore, the guidance should also consider how to demonstrate where models are intended to be used for activities and processes currently performed by humans. The models are adequately developed and protected and certified—even if these processes are not explicitly certified due to assumptions under personnel licensing or other considerations.

The AIA Civil Aviation Cybersecurity Subcommittee recommends that efforts between SAE G-34, EUROCAE WG-114, and the intended new Special Committee within RTCA are coordinated to avoid duplication, redundancy, and conflict between developing standards for AI/ML use in aviation. One proposal is for G-34 and EUROCAE WG-114 to develop standards for the processes for developing, deploying, and either certifying or approving AI/ML for aviation. It is also recommended that RTCA explore standards for specific implementation or technologies of AI/ML as well as where existing standards could be augmented by AI/ML.

6 Abbreviations

AC	Advisory Circular
AI	Artificial Intelligence
AIA	Aerospace Industries Association
AIAA	American Institute of Aeronautics and Astronautics
AL	Assurance Level
AMC	Acceptable Means of Compliance
ATM	Air Traffic Management
CNS	Communication, Navigation, Surveillance
CS	Certification Specifications
DAH	Design Approval Holder
DAL	Design / Development Assurance Level
EASA	European Union Aviation Safety Agency
EO	Executive Order
EUROCAE	European Organisation for Civil Aviation Equipment
FAA	Federal Aviation Administration
INCS	Integrated Communication, Navigation and Surveillance Conference
IT	Information Technology
LLM	Large Language Model
MASP	Minimum Aviation System Performance Standards
MOPS	Minimum Operational Performance Standards
ML	Machine Learning
NCSC	National Cyber Security Centre
NIST	National Institute of Standards and Technology

NSF	National Science Foundation
OT	Operational Technology
OWASP	Open Worldwide Application Security Project
SAL	Security Assurance Level
SciTech	Science and Technology Forum
SDO	Standards Development Organization
SRMA	Sector Risk Management Agency
TC	Type Certificate
WG	Working Group

7 List of references

Reference	**Title**
AIAA SciTech 2024 [Wes Ryan, Northrop Grumman Corporation]	Using Machine Learning in Safety-Critical Systems: Regulator and Technologist Perspectives: American Institute of Aeronautics and Astronautics Science and Technology Forum 2024
EASA Artificial Intelligence concept paper	EASA Concept Paper: First usable guidance for Level 1 machine learning applications
EASA Artificial Intelligence concept paper (draft Issue 2)	EASA Concept Paper: First usable guidance for Level 1&2 machine learning applications
EASA Artificial Intelligence Roadmap 2.0	Artificial Intelligence Roadmap 2.0: Human-centric approach to AI in aviation
EASA CS-25	Certification Specifications for Large Aeroplanes
EASA ED 2020/006/R	EASA Executive Director Decision Aircraft Cybersecurity
EU 2022/1645	Commission Delegated Regulation (EU) 2022/1645
EU 2023/203	Commission Implementing Regulation (EU) 2023/203
EUROCAE ED-109A *(equivalent to DO-278A)*	Software Integrity Assurance Considerations for Communication, Navigation, Surveillance and Air Traffic Management (CNS/ATM) Systems

Reference	Title
EUROCAE ED-12C *(equivalent to DO-178C)*	Software Considerations in Airborne Systems and Equipment Certification
EUROCAE ED-153	Guidelines for ANS Software Safety Assurance
EUROCAE ED-202A *(equivalent to DO-326A)*	Airworthiness Security Process Specification
EUROCAE ED-203A *(equivalent to DO-356A)*	Airworthiness Security Methods and Considerations
EUROCAE ED-206 *(equivalent to DO-392)*	Guidance on Security Event Management
EUROCAE ED-215 *(equivalent to DO-330)*	Software Tool Qualification Considerations
FAA AI Roadmap (Draft 0.2)	Roadmap for Artificial Intelligence Safety Assurance
EO 13960	Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence
ICNS 2022	Defining an Initial Classification Scheme for Non-Deterministic AI Technologies: Integrated Communication, Navigation and Surveillance Conference 2022
Microsoft Vulnerability Severity Classification for AI Systems	https://www.microsoft.com/en-US/msrc/aibugbar
NCSC: ChatGPT and large language models: what's the risk	https://www.ncsc.gov.uk/blog-post/chatgpt-and-large-language-models-whats-the-risk
NCSC: Guidelines for secure AI system development	https://www.ncsc.gov.uk/collection/guidelines-secure-ai-system-development
NCSC: Thinking about the security of AI systems	https://www.ncsc.gov.uk/blog-post/thinking-about-security-ai-systems
OWASP Machine Learning Top Ten	https://owasp.org/www-project-machine-learning-security-top-10/
OWASP Top 10 for LLM 2023	https://owasp.org/www-project-top-10-for-large-language-model-applications/assets/PDF/OWASP-Top-10-for-LLMs-2023-v05.pdf

Reference	**Title**
PS-AIR-21.16-02	Establishment of Special Conditions for Cyber Security
RTCA DO-178C *(equivalent to ED-12C)*	Software Considerations in Airborne Systems and Equipment Certification
RTCA DO-278A *(equivalent to ED-109A)*	Software Integrity Assurance Considerations for Communication, Navigation, Surveillance and Air Traffic Management (CNS/ATM) Systems
RTCA DO-326A *(equivalent to ED-202A)*	Airworthiness Security Process Specification
RTCA DO-330 *(equivalent to ED-215)*	Software Tool Qualification Considerations
RTCA DO-356A *(equivalent to ED-203A)*	Airworthiness Security Methods and Considerations
RTCA DO-392 *(equivalent to ED-206)*	Guidance on Security Event Management