2021 AAAI CONFERENCE: AI SAFETY WORKSHOP

AI for Future Skies: On-going standardization activities to build the next certification/approval framework for airborne and ground aeronautical products

Authors:

Christophe GABREAU, Airbus, Co-chair of EUROCAE WG-114 Group

Beatrice PESQUET-POPESCU, Thales LAS/AMS, Co-chair of EUROCAE WG-114 Group

Fateh KAAKAI, Thales LAS/AMS, Sub-Group Leader of EUROCAE WG-114 Group

Baptiste Lefeuvre, Thales AVS, Sub-Group Leader of EUROCAE WG-114 Group

A joint group EUROCAE WG-114 / SAE G-34 (AI in Aviation)

500+ engineers

Researchers and AI scientists from across the globe, with representation from regulators and authorities (FAA, EASA, TCCA, ANAC, EDA, NASA, DOD, EUROCONTROL), major airframers, UAS/UAM/eVTOL manufacturers, engine manufacturers, component manufacturers, technology providers, and other stakeholders, including operators and airlines

Special thanks to all contributors

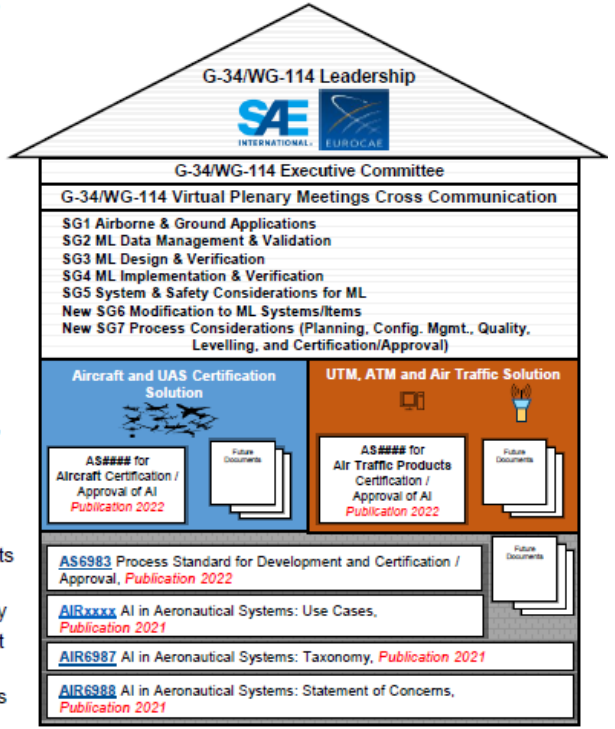
G-34/WG-114 focuses on implementation and certification related to AI technologies for the safer operation of aerospace systems and aerospace vehicles.

G-34/WG-114 (comprised of 500+ members) promotes and standardizes Artificial Intelligence in the entire aviation eco-system (both Airborne and Ground) addressing both manned and UAS.

G-34/WG-114's Global contributors: Boeing, Airbus, ATR, Embraer, Textron, Gulfstream, Dassault, Mitsubishi, Lockheed, Northrop Grumman, GA-ASI, HondaJet, Daher, IAI, ICAO, FAA, EASA, TCCA, ANAC, DGAC, CAA UK, CAA NZ, JCAB, ENAC, FOCA, DOD, EDA, Lilium, Aerion Supersonic, Amazon, DXC, SAP, IBM, Joby, EUROCONTROL, NASA, EDA, Honeywell, Collins, Thales, GE, P&W, RR, Safran, Raytheon, BAE, Elbit, L3Harris, Iridium, Japan Manned Space Systems, FedEx, UPS, AF-KLM, Nodein, Lufthansa, Audi, Toyota, IATA, Leonardo, Leidos, NVIDIA, Intel, Saab, Volocopter, ANSPs, Skyguide, Searidge, Woodward, Vertical Aerospace, Diehl, ADB Safegate, AVSI, ANSYS, BNAE, Copenhagen Airports, D-Risq, Daedalean AI, KIAST, Infosys, Afuzion, Patmos Engineering, QinetiQ, RelmaTech, Rockdale Systems, DLR, drR2, Federated Safety, MathWorks, SRI, Oak Ridge National Lab, etc.

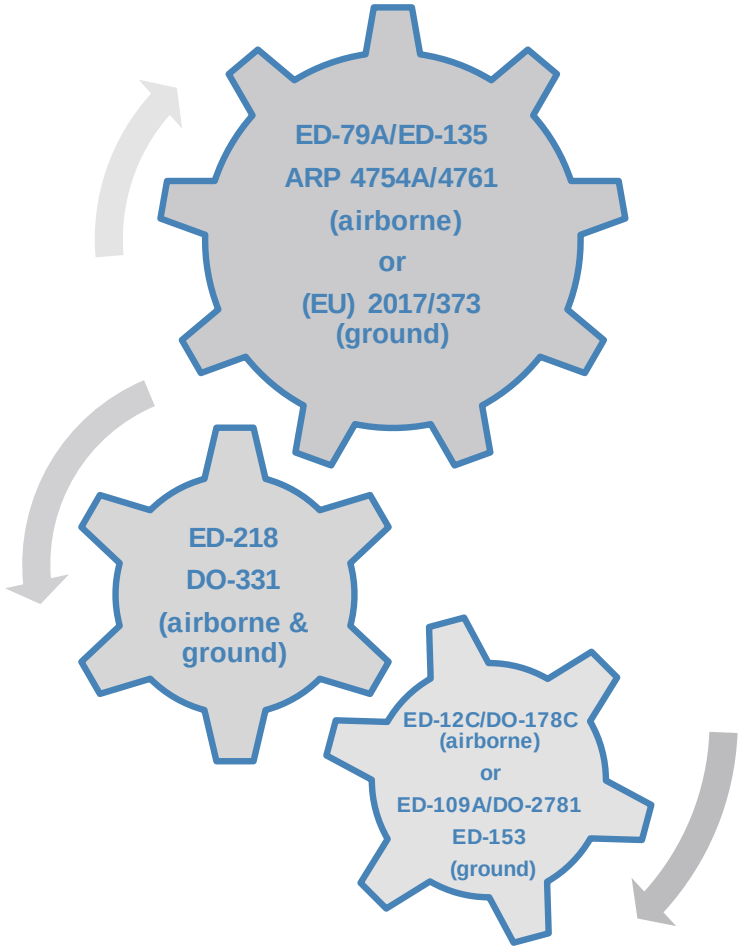
Works In Progress and deliverables:

- AS6983 Process Standard for Development and Certification/Approval of Aeronautical Safety-Related Products Implementing AI
- AIR6987 Artificial Intelligence in Aeronautical Systems: Taxonomy
- AIR6988 Artificial Intelligence in Aeronautical Systems: Statement of Concerns
- AIRxxxx Artificial Intelligence in Aeronautical Systems: Use Cases Considerations



For more information and/or membership registration, contact: jordanna.bucciere@sae.org and/or anna.guegan@eurocae.net.

Existing standardization framework



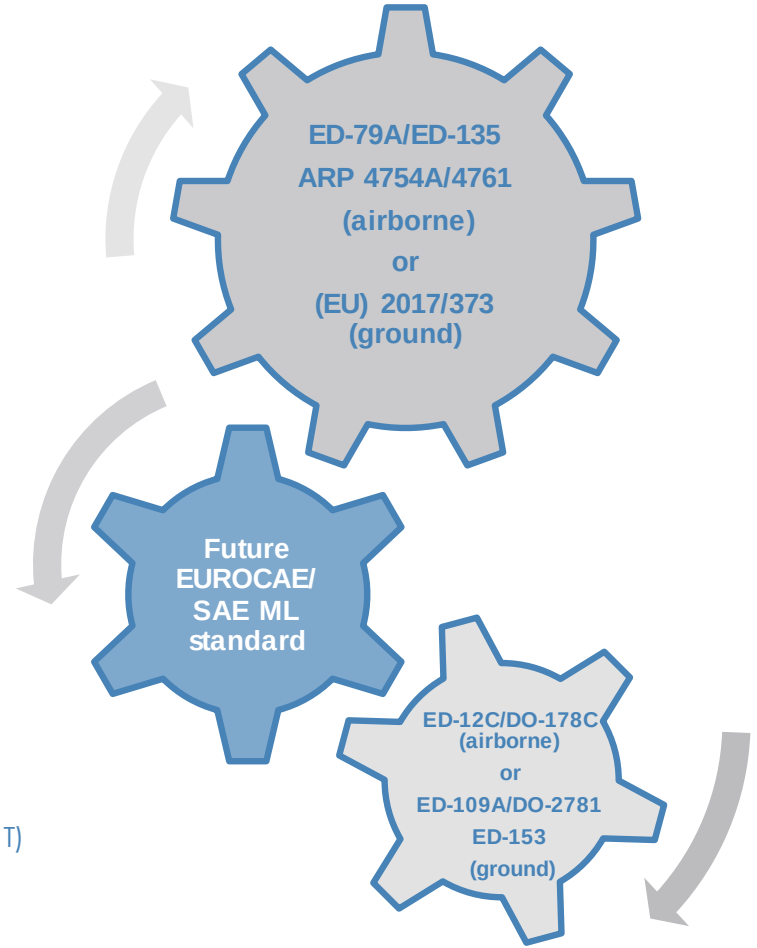
SYSTEME

MODEL

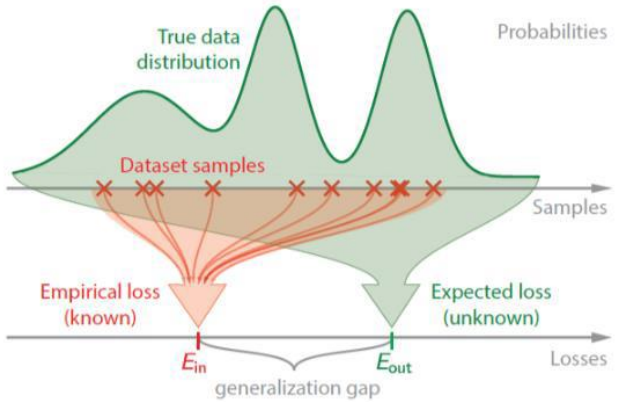
ITEM
(COMPONENT)

Future standardization framework

WIP



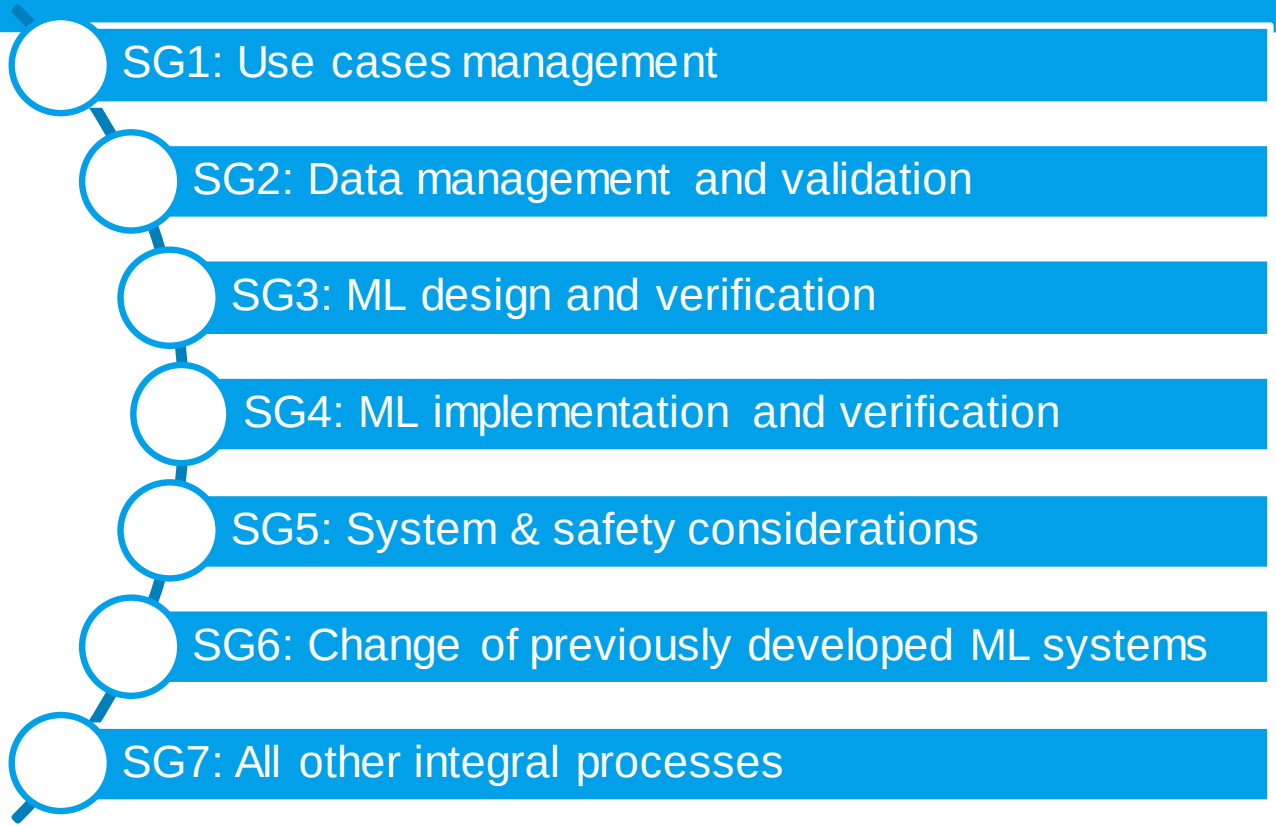
Robustness and Verification Challenges



- Evaluate robustness of an algorithm to changes in the training set
- Detect unintended and unexpected behavior of NN
- Detect abnormal or adversarial inputs to the NN
- Asses intrinsic robustness of trained ML models through formal or empirical methods
- Assess training methodologies that can enhance or guarantee robustness
- Manage performance / robustness tradeoff
- Define safety process analysis and relevant architectural mitigations (bounding, voting, diversity, etc)

Source: EASA CodaNN IPC report

WG-114/G-34 setup for the standard organization



SG3: Scope and desired attributes of the MLDL

The MLDL should be:

Generic

- The MLDL is applicable to offline ML technologies considered in G34-WG114 scope
- Any technology-specific MLDL phase should be addressed as a second step (further updates of the MLDL)

Process/Environment Agnostic

- The MLDL does not impose a specific development process
- The MLDL does not impose a specific learning environment

Support certification/approval

- ML assurance objectives should be well organized consistently with MLDL
- ML assurance objectives should be simple and clear

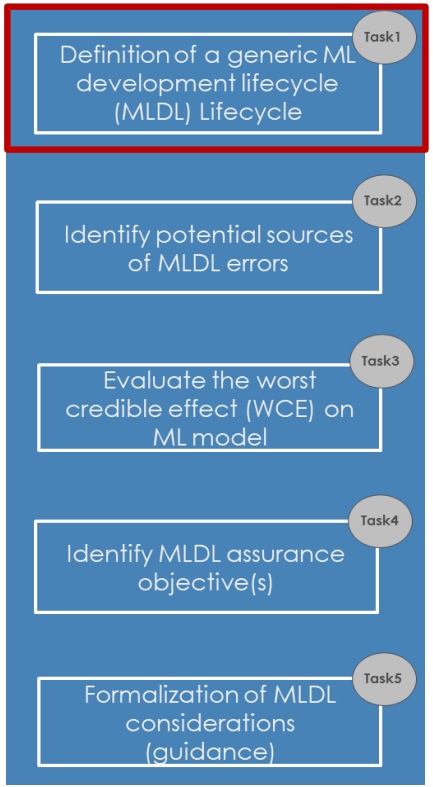
Counter-examples

e.g. The MLDL is only applicable to supervised learning using Artificial Neural Network

- e.g. The MLDL is only applicable for V or W development process
- e.g. The MLDL is only applicable to ML models built using tensorflow framework

- e.g. ML assurance objectives are organized using phases and steps that are not consistent with the MLDL definition

Methodology to build Machine Learning Assurance Objectives



Task 1 Objective:
Define a generic ML development lifecycle (MLDL) to support:
- the analysis of fault injection all along the ML developement lifecycle
- the identification of ML development assurance objectives (MLDAO) to avoid fault injection or detect resulting errors
- the evaluation of proposed MLDAO with appropriate use cases.

This MLDL should be approved by the full SG3 group

Task 2 Objective:
Identify the possible source of errors called either ML development fault injection cases or ML development failure modes. They are described with at least the following attributes : Name, Rationale (if not obvious)
The completness of the failure modes should be assessed using appropriate method(s).
The list of MLDL failure modes can be classified per MLDL phase and should be approved by the full SG3 group

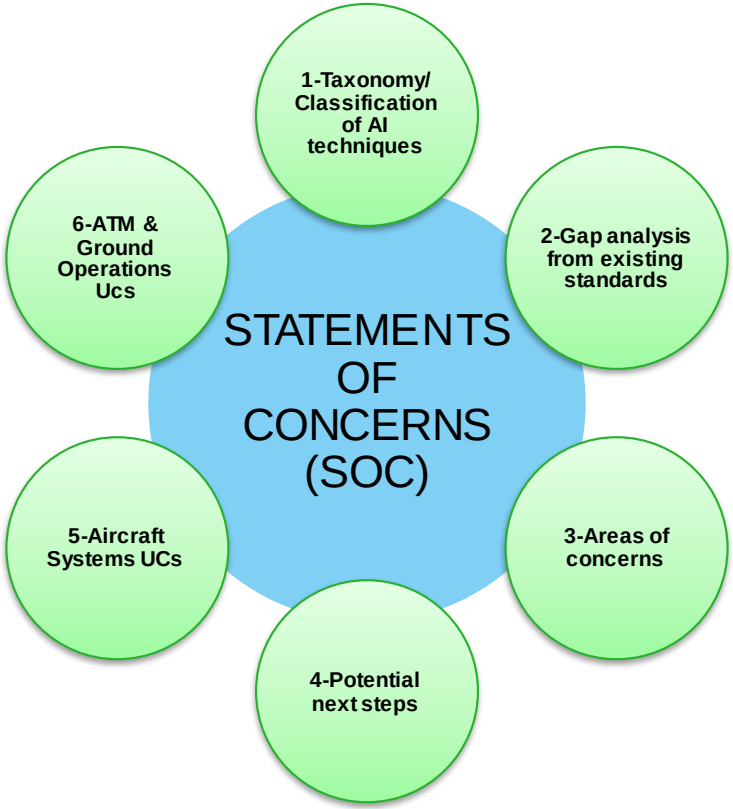
Task 3 Objective:
Study the worst credible effect (WCE) on the ML model of all ML development failure modes. The adverserial effects that are considered to establish WCEs come from SG5 safety objectives (e.g. impact on ML model integrity, performance, explainability, etc.). When not obvious, a rationale should be provided to explain WCEs. When there is no adverserial effect on safety, the WCE should be « No identified effect ».

Task 4 Objective:
Identify MLDL assurance objective(s) to mitigate any adverserial WCE on SG5 safety objectives allocated to the ML model (e.g. adverserial impact on ML model integrity, performance, explainability, etc.). MLDL assurance objectives should be classified by DAL/AL/SWAL levels. A gradation of these assurance objectives is expected according to the DAL/AL/SWAL levels. Airborne and Ground specificities should be taken into account. A rationale should be provided to explain each MLDL assurance objective. When there is no adverserial effect on safety (i.e., WCE = No identified effect »), no assurance objective is needed.

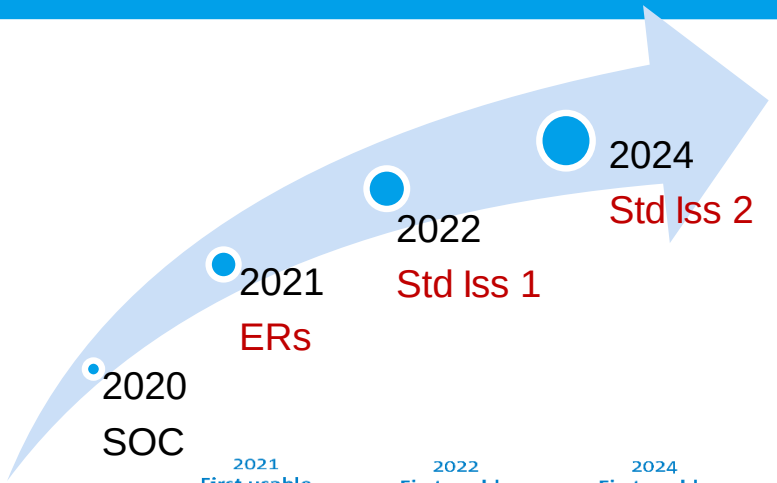
Task 5 Objective:
Formalize the outputs of all tasks into a guidance material that follows AS6983/ED-XXX Outline. This guidance is expected to be part of the final AS6983/ED-XXX standard. The need to issue a FAQ should be assessed by SG3 leaders.

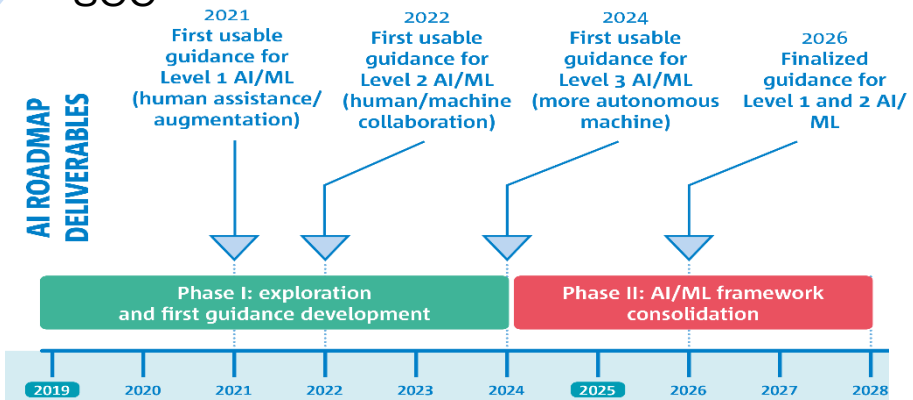
2021 Outcomes: Statement of Concerns

Worldwide industries aligned on the same concerns



WG-114/G-34 Roadmap





Deliveries

- ✓ SOC (Statement of Concerns) – ER/AIR
- ✓ Taxonomy, Use Cases – ER/AIR
- ✓ Std Issue 1: ML (Offline Learning) – ED/AS
- ✓ Std Issue 2: Other AI Technologies – ED/AS

EASA Roadmap

THANK YOU FOR YOUR ATTENTION !

Questions ?

Source: https://xkcd.com/