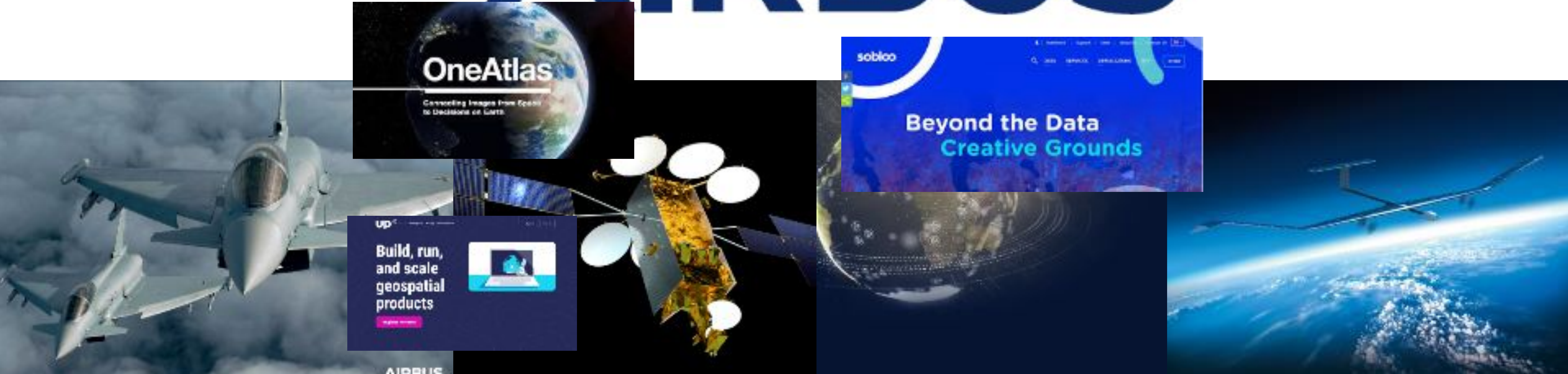
FIIA 2021 Trustworthy Artificial Intelligence in AIRBUS

Romaric Redon AIRBUS
Head Advisor Artificial Intelligence

AIRBUS

skywise.
With Skywise, unleash the full potential of every aircraft

AIRBUS


OneAtlas
Connecting Images from Space to Decisions on Earth
sobloo
Beyond the Data
Creative Grounds
up42
Build, run, and scale geospatial products
AIRBUS

What do we do with Artificial Intelligence @AIRBUS

Perceive / Observe

Computer Vision

Pattern recognition and Times Series Analysis

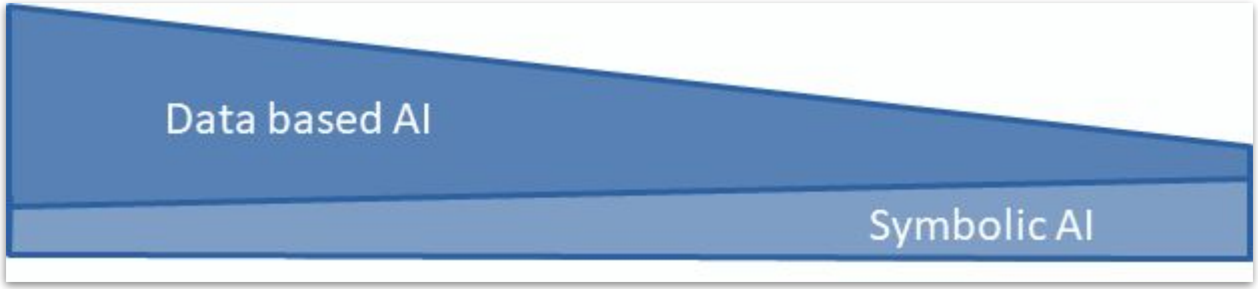
Natural Language understanding and Processing

Predict / forecast / Orient

Hybrid Modelling

Decide/Act

Decision making



AIRBUS

What do we do with Artificial Intelligence @AIRBUS

Perceive / Observe

Computer Vision TRUST

Pattern recognition and Times Series Analysis TRUST

Natural Language understanding and Processing TRUST

TRUST
for all application cases
for all AI technologies

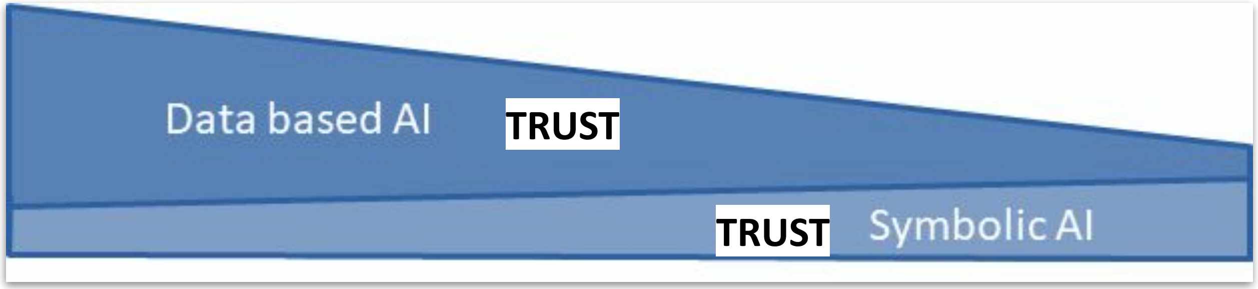
TRUST is ensured at
System level

Predict / forecast / Orient

Hybrid Modelling TRUST

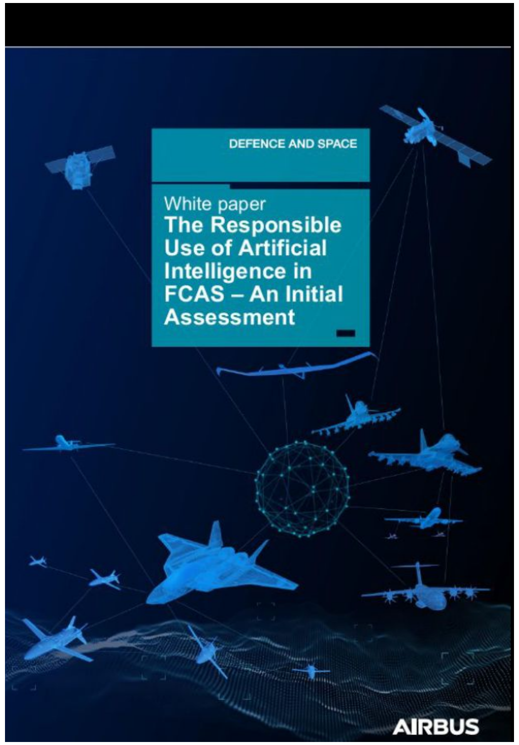
Decide/Act

Decision making TRUST



Trustworthy AI - Several dimensions

Responsible use of AI

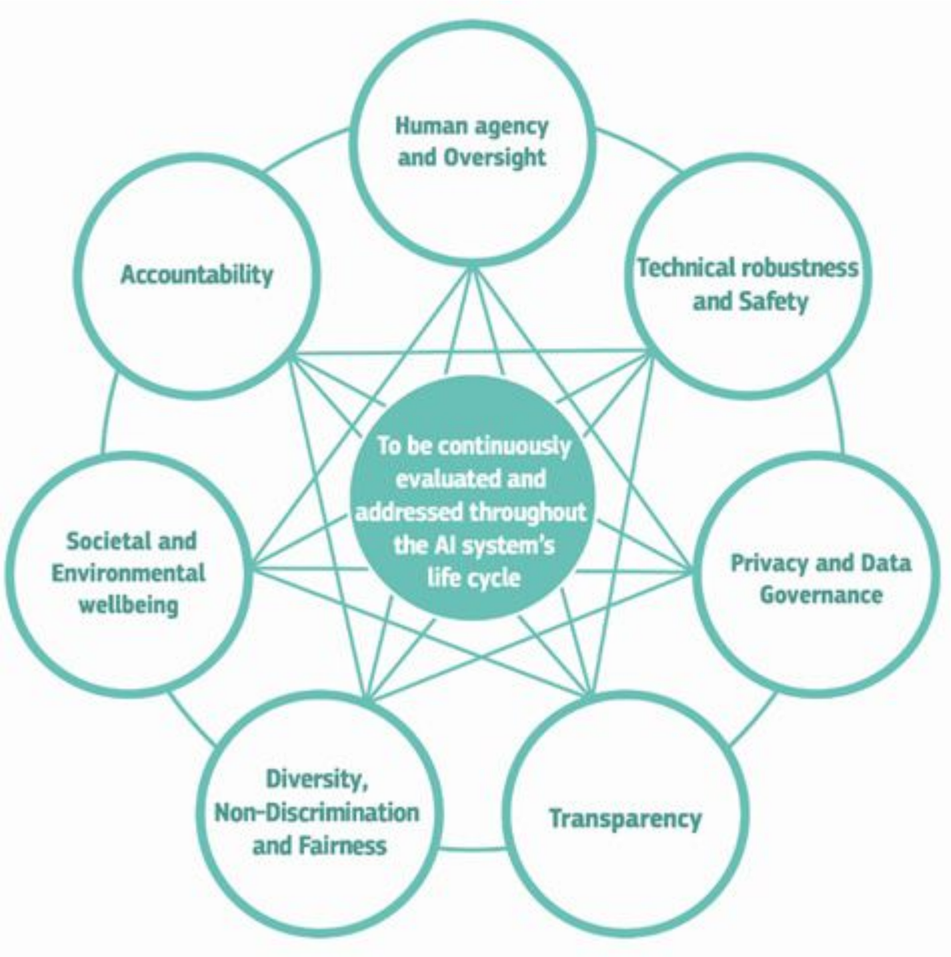


DEFENCE AND SPACE

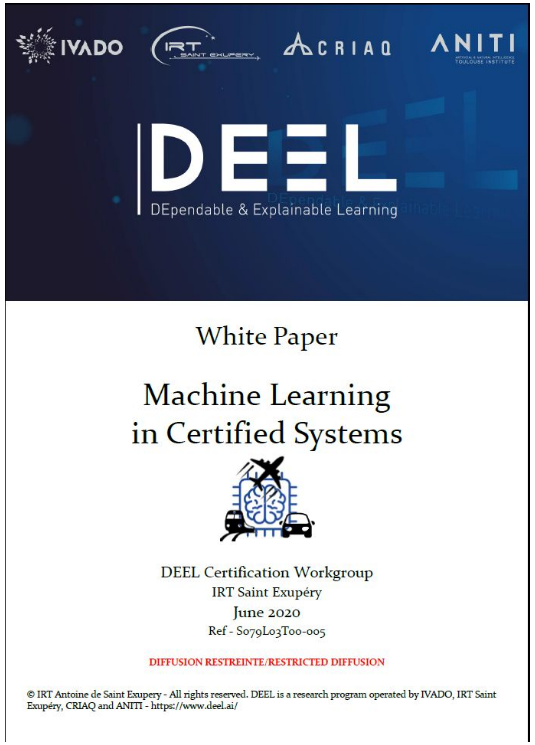
White paper
The Responsible Use of Artificial Intelligence in FCAS – An Initial Assessment

AIRBUS

http://www.fcas-forum.eu/en


Human agency and Oversight
Technical robustness and Safety
Privacy and Data Governance
Transparency
Diversity, Non-Discrimination and Fairness
Societal and Environmental wellbeing
Accountability
To be continuously evaluated and addressed throughout the AI system's life cycle

Safe use of AI



IVADO IRT SAINT EXUPERY CRIAQ ANITI

DEEL
DEpendable & Explainable Learning

White Paper

Machine Learning in Certified Systems

DEEL Certification Workgroup
IRT Saint Exupéry
June 2020
Ref - S079L03T00-005

DIFFUSION RESTREINTE/RESTRICTED DIFFUSION

© IRT Antoine de Saint Exupery - All rights reserved. DEEL is a research program operated by IVADO, IRT Saint Exupéry, CRIAQ and ANITI - https://www.deel.ai/

https://arxiv.org/abs/2103.10529

Fairness

Robustness

Explainability

Trustworthy AI requirements vs criticality levels

NOT CRITICAL

Advanced search for Engineering documentation

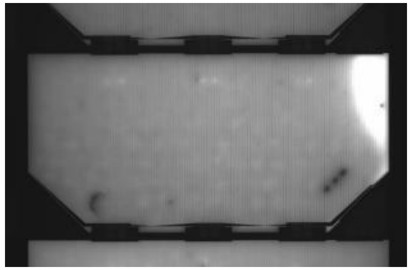
AI for Human resources applications

AI

HIGH RISK

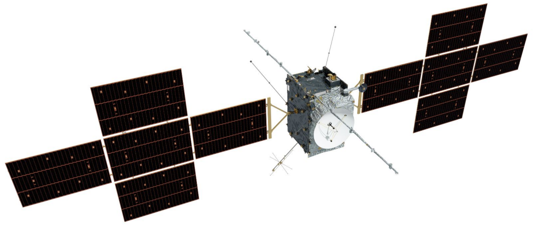
BUSINESS CRITICAL

Visual inspection of solar Panels



MISSION CRITICAL

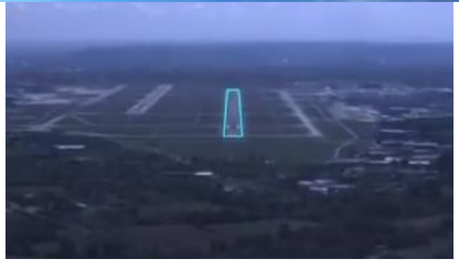
Guidance and navigation of spacecrafts



ECSS

SAFETY CRITICAL

Assistance to pilots



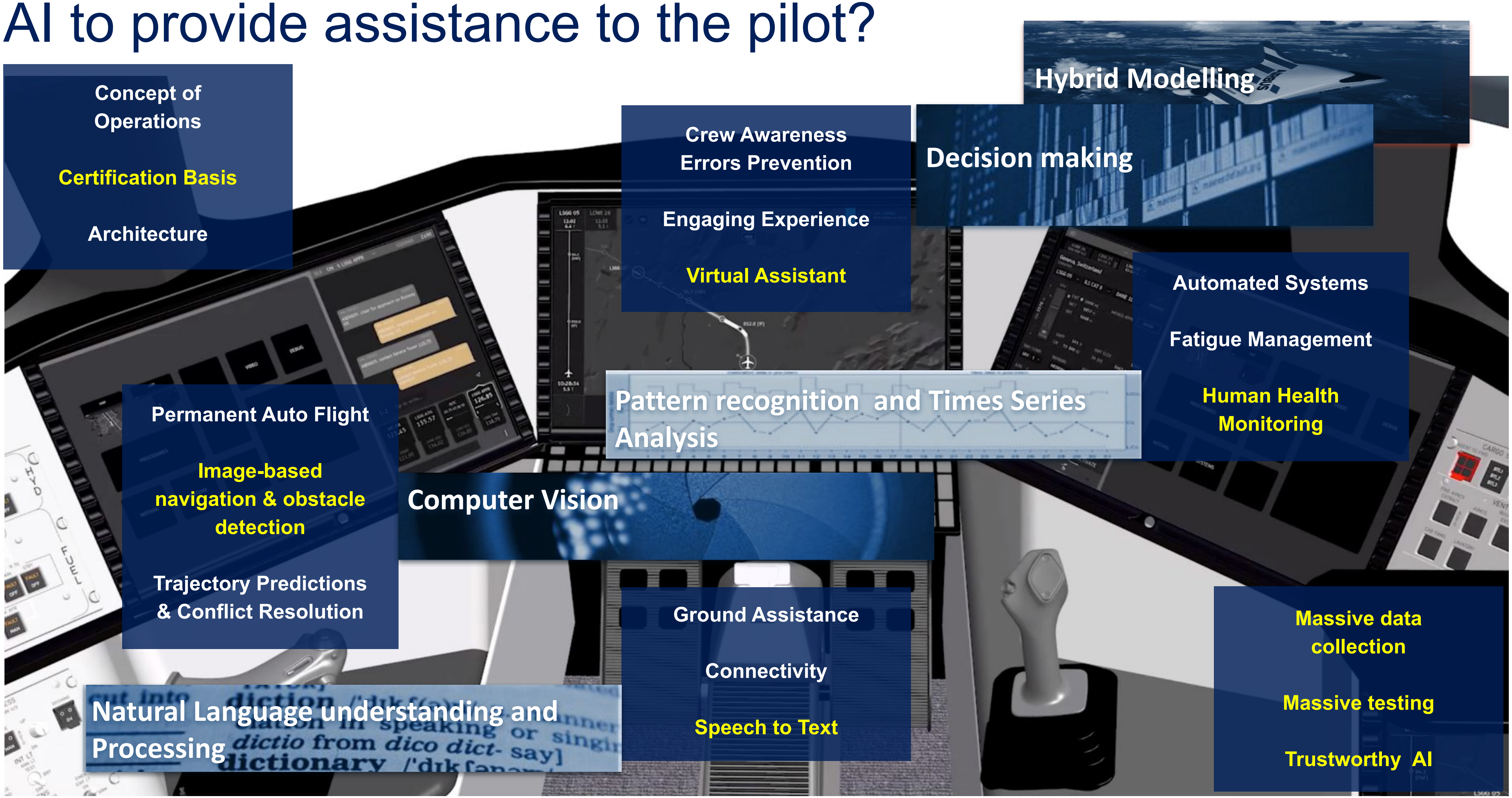
EASA
European Union Aviation Safety Agency

AI

HIGH RISK

AIRBUS

AI to provide assistance to the pilot?



Concept of Operations

Certification Basis

Architecture

Crew Awareness Errors Prevention

Engaging Experience

Virtual Assistant

Hybrid Modelling

Decision making

Automated Systems

Fatigue Management

Human Health Monitoring

Pattern recognition and Times Series Analysis

Permanent Auto Flight

Image-based navigation & obstacle detection

Trajectory Predictions & Conflict Resolution

Computer Vision

Ground Assistance

Connectivity

Speech to Text

Massive data collection

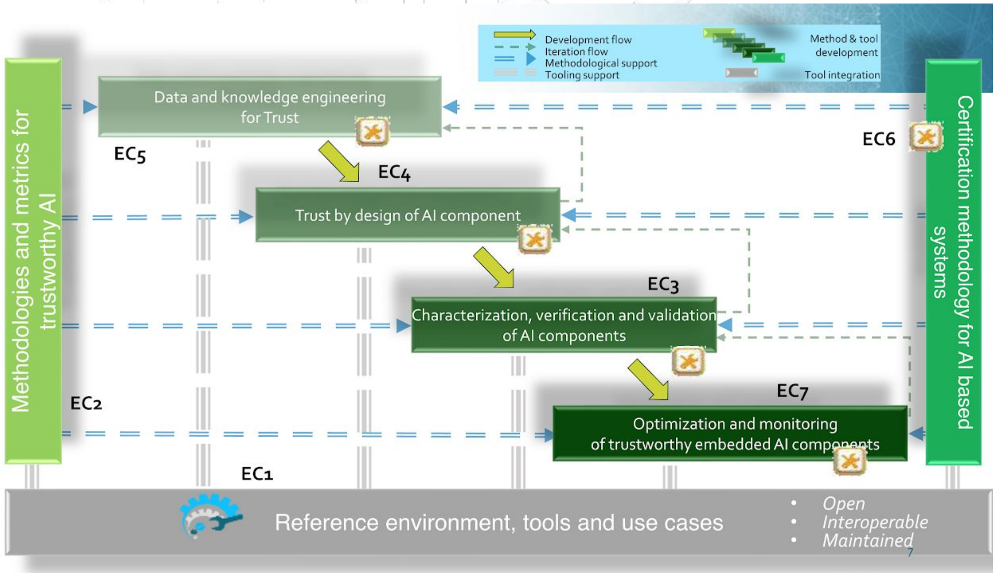
Massive testing

Trustworthy AI

Natural Language understanding and Processing

Trustworthy AI Engineering

- Requirement and Data Eng for Trustworthy AI
- Design for Trustworthy AI
- Monitoring and fail safe Architectures
- Trust On-board HW for AI & code generation
- V&V methods for critical systems with AI
- AI regulation and Standards - certification


Development flow
Iteration flow
Methodological support
Tooling support
Method & tool development
Tool integration
Methodologies and metrics for trustworthy AI
Data and knowledge engineering for Trust
EC5
EC4
Trust by design of AI component
EC3
Characterization, verification and validation of AI components
EC7
Optimization and monitoring of trustworthy embedded AI components
EC2
EC1
EC6
Certification methodology for AI based systems
Reference environment, tools and use cases
Open
Interoperable
Maintained

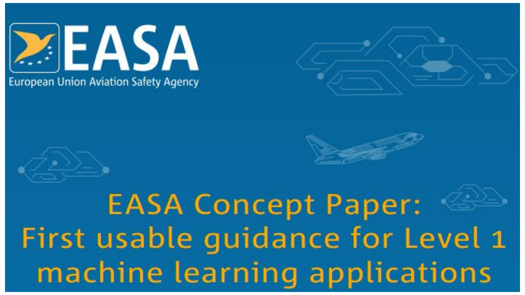
AIRBUS

Requirement and Data Eng for Trustworthy AI

With Data Driven AI Spec ~ Data

Difficulty to define the Operational Design Domain for high dimensional input space

How to specify the data needed to cover the ODD?
How to detect undesired biais?
How to choose the right distribution?
How to use a good mix of real/synthetic data to have the right distribution?



EASA
European Union Aviation Safety Agency

EASA Concept Paper:
First usable guidance for Level 1 machine learning applications

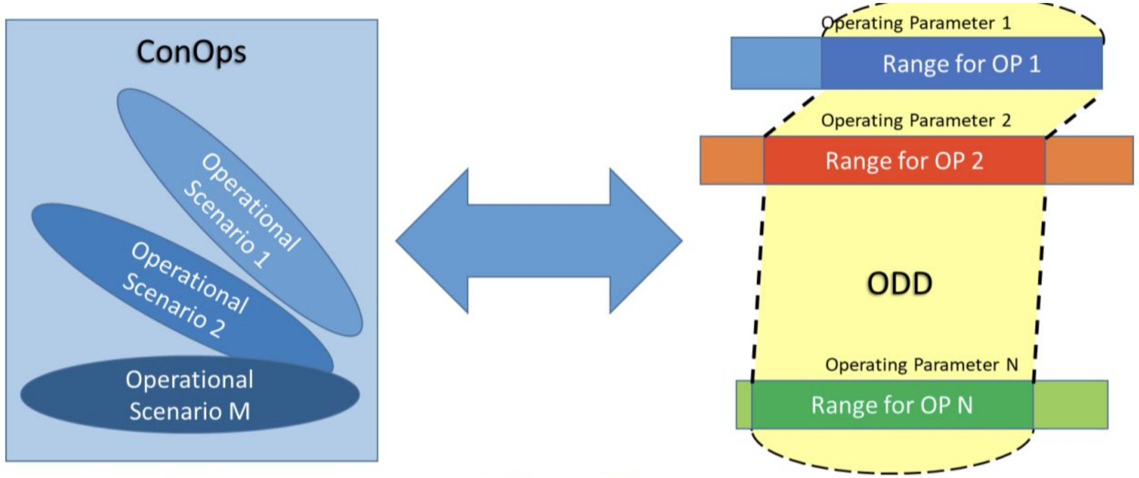


Figure 5 — Interrelationship between ConOps and ODD

Objective CO-03: The applicant should define and document the ConOpsfor all AI-based (sub)systems. A focus should be put on the definition of the **operational design domain (ODD)** and on the capture of specific operational limitations and assumptions.

Design for Trustworthy AI - Robustness

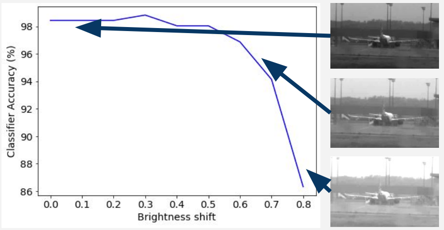
How to improve robustness ?

Generalisation guarantees : how accurately a Machine Learning algorithm is able to predict outcome values for previously unseen data?

Natural perturbations

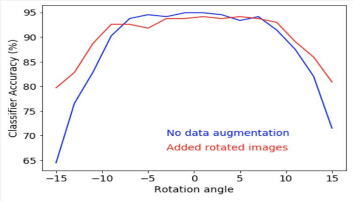
1) Robustness assessment

Performance evaluation under artificial corruptions, e.g. brightness change, rotations, occlusions, rain, etc.



2) Data augmentation

Use of perturbed images during training improves the model robustness

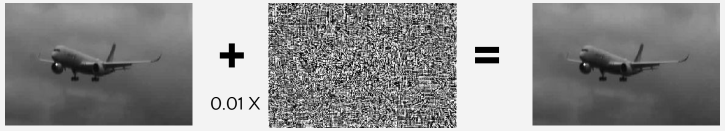


Adversarial attacks

IBM ART

1) Attack benchmarks

Up to 96% of all inputs can be successfully attacked (i.e. visually indistinguishable but wrong model prediction)



Aircraft

Perturbation

Not an aircraft

2) Adversarial defenses

- **Adversarial training**: attacks in training reduces success rate from 96% to 34%
- **Smoothing defense**: trade-off between classification accuracy and robustness

Original image

Attacked image used in training

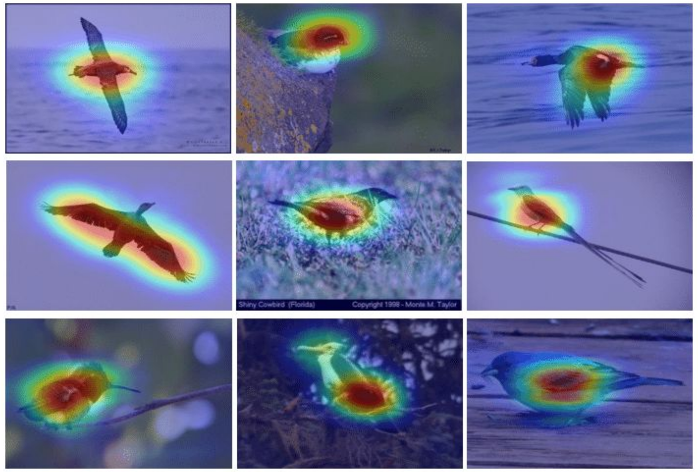
Design for Trustworthy AI - Explicability

“The AI explainability deals with the capability to provide the human with understandable and relevant information on how an AI/ML application is coming to its results.” EASA guidelines level 1

Explainability for

- ML model designer/authorities
- Operator
- forensic investigations

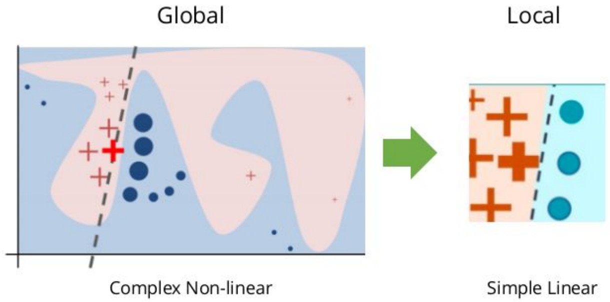
Saliency Maps



He, Xiangteng & Peng, Yuxin & Zhao, Junjie. (2017). Fine-grained Discriminative Localization via Saliency-guided Faster R-CNN.

ML interpretation models

SHAP,LIME,...



How good is your Explanation?
How could you use it to provide guarantees?

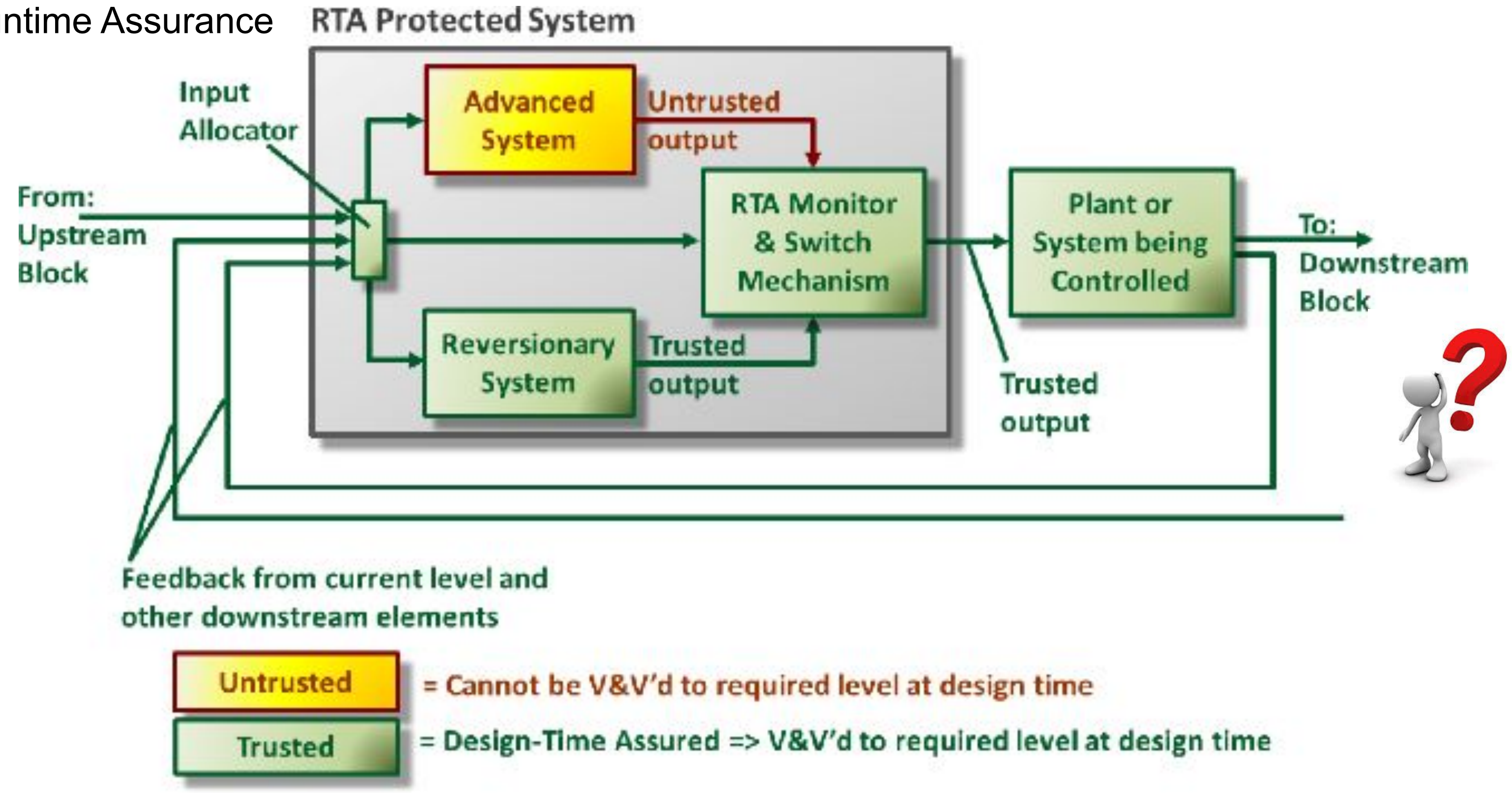
What about NLP and Speech, Planning & Scheduling?

AIRBUS

Monitoring and fail safe architecture

- ODD Monitoring verifies that the ML-based system is operated in its usage domain
- OOD Monitoring ensures that the ML Model operates in the distribution defined during the training process.
- Attacks monitoring allows to detect adversarial attacks.
- Robustness monitoring ensures that the ML Model is used in a stable area.
- Consistency monitoring analyzes the consistency of outputs.

Runtime Assurance



Untrusted = Cannot be V&V'd to required level at design time

Trusted = Design-Time Assured => V&V'd to required level at design time

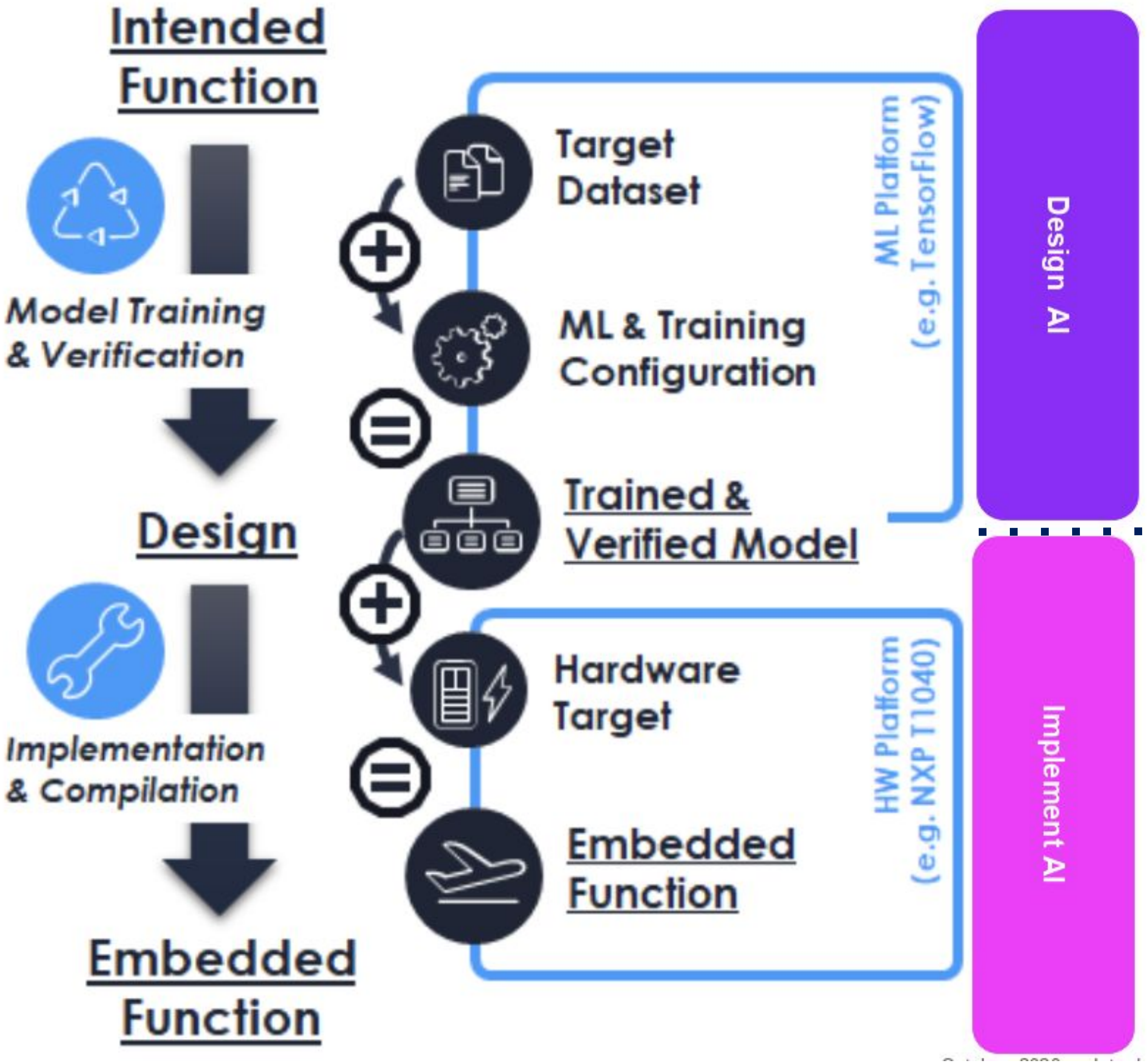
What is the more efficient monitoring architecture?

What would be a Dissimilar Architecture with ML?
Is Same data set - different models enough ?

How to find the best trade-of
Impact of monitoring on availability?

Schierman, John D. et al. “Runtime Assurance Framework Development for Highly Adaptive Flight Control Systems.” (2015).

Trusted on-board hardware for AI + code generation

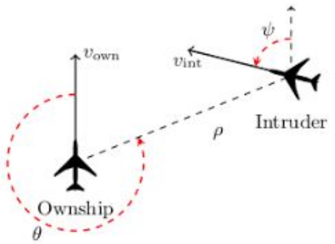

Intended Function
Model Training & Verification
Design
Implementation & Compilation
Embedded Function
Target Dataset
ML & Training Configuration
Trained & Verified Model
Hardware Target
Embedded Function
ML Platform (e.g. TensorFlow)
HW Platform (e.g. NXP T1040)
Design AI
Implement AI

MODEL FREEZE

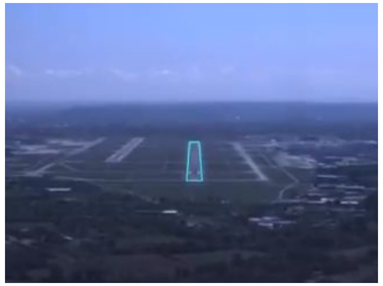
How to generate efficient Embedded code while preserving the semantic of the model?

Which High Performance Computing and Certifiable on-board Hardware can be used?

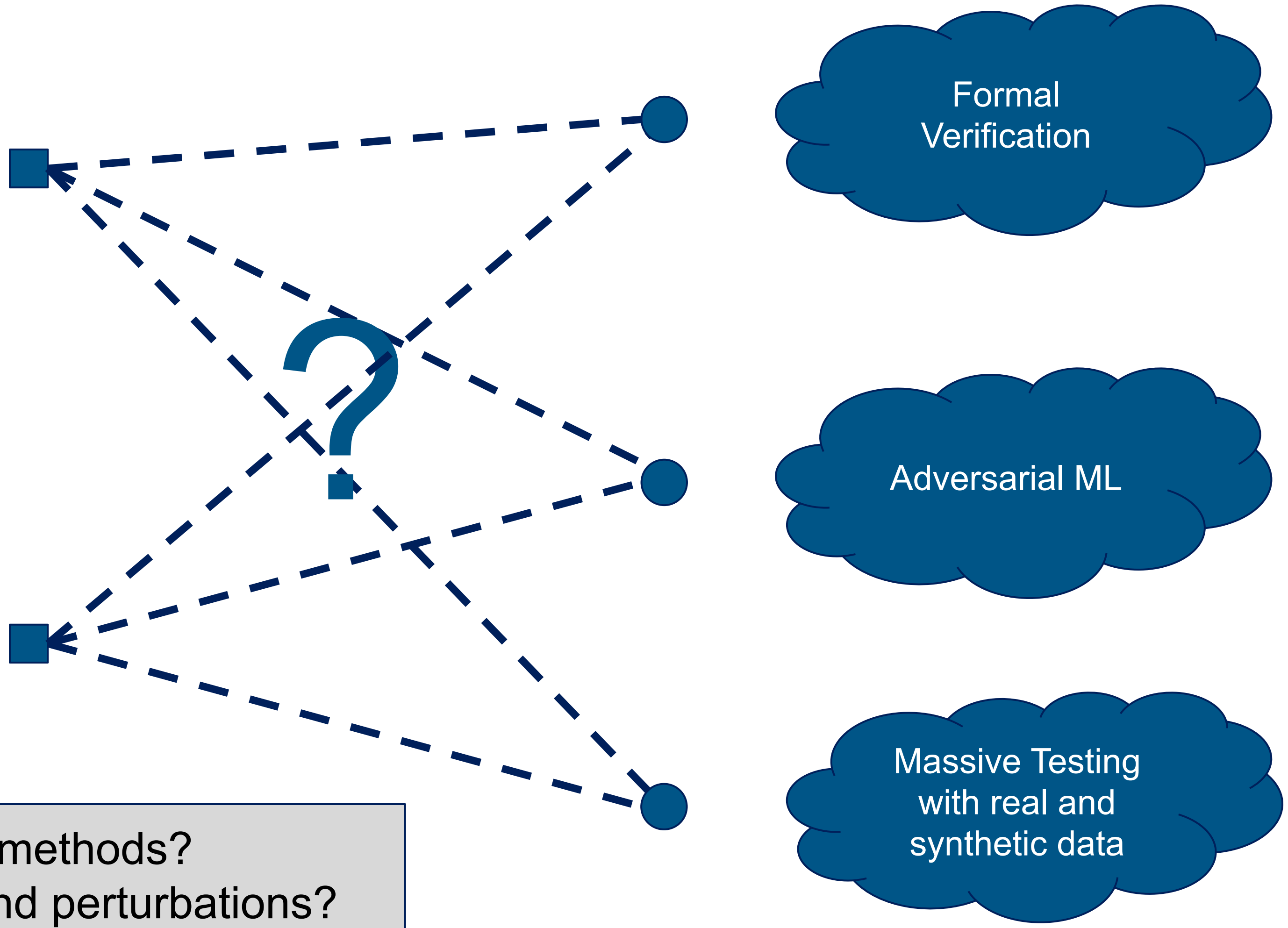
V&V methods for critical systems with AI



NN to approximate complex functions (table, data, physical models)



Deep NN for Vision Based Navigation



How to improve scalability of formal methods?
How to define the good properties and perturbations?
How to perform more efficient/frugal massive Testing?
How to qualify synthetic data generator?
How to combine/Hybridize methods?

AI regulation and Standards

SECTORIAL

WG-114 EUROCAE

EUROPEAN COOPERATION ECSS FOR SPACE STANDARDIZATION

DGA

BUNDESWEHR

Ministry of Defence

GOBIERNO DE ESPAÑA MINISTERIO DE DEFENSA

■ ■ ■

How General and Sectorial regulations will be articulated ?

GENERAL

AI

ISO

■ ■ ■

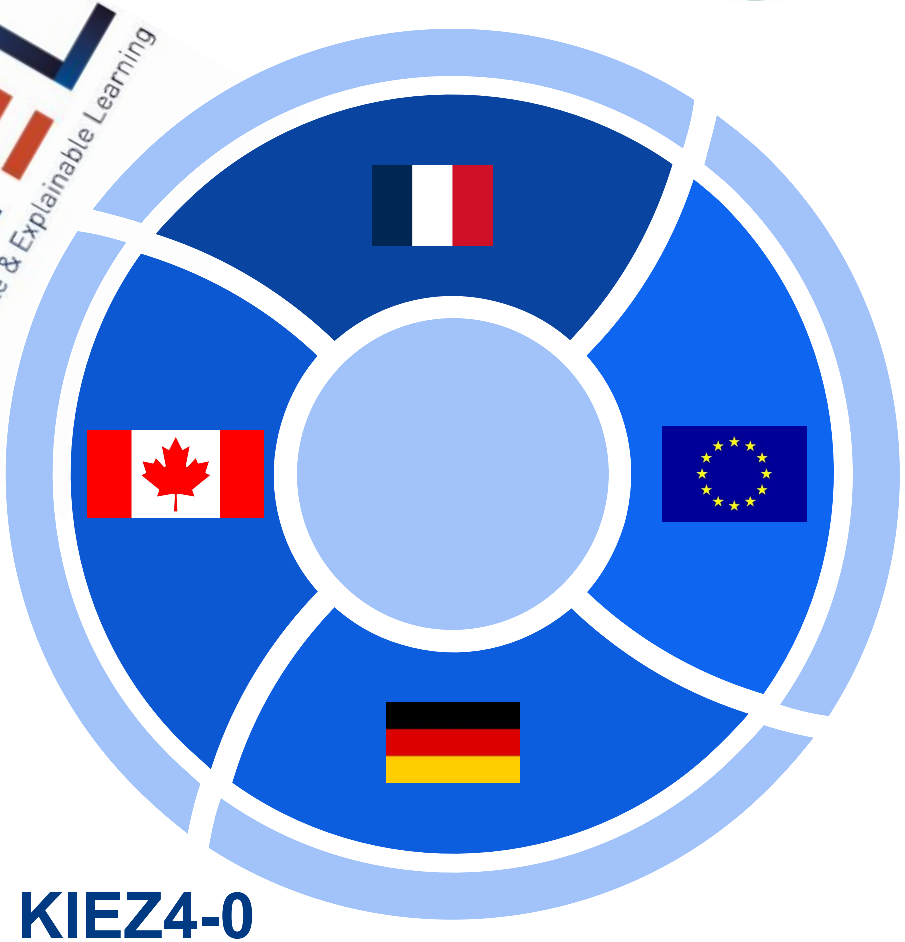
AIRBUS

Collaborative Research project about certifiable and trusted AI

ANITI
ARTIFICIAL & NATURAL INTELLIGENCE
TOULOUSE INSTITUTE

Confiance ai

DEEL
DEpendable & Explainable Learning



HUMANE AI NET

INDEPENDENT
HIGH-LEVEL EXPERT GROUP ON
ARTIFICIAL INTELLIGENCE
SET UP BY THE EUROPEAN COMMISSION
AI

TAILOR

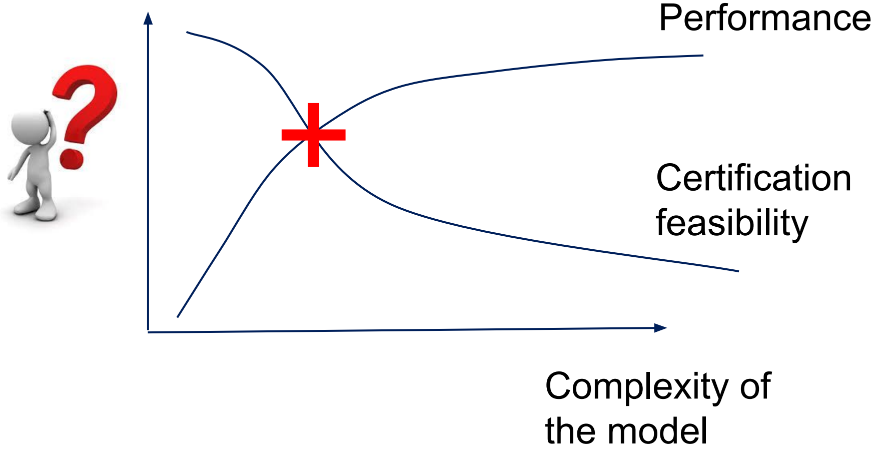
KIEZ4-0

Künstliche Intelligenz Europäische Zertifizierung unter Industrie 4.0 Aspekten

AIRBUS

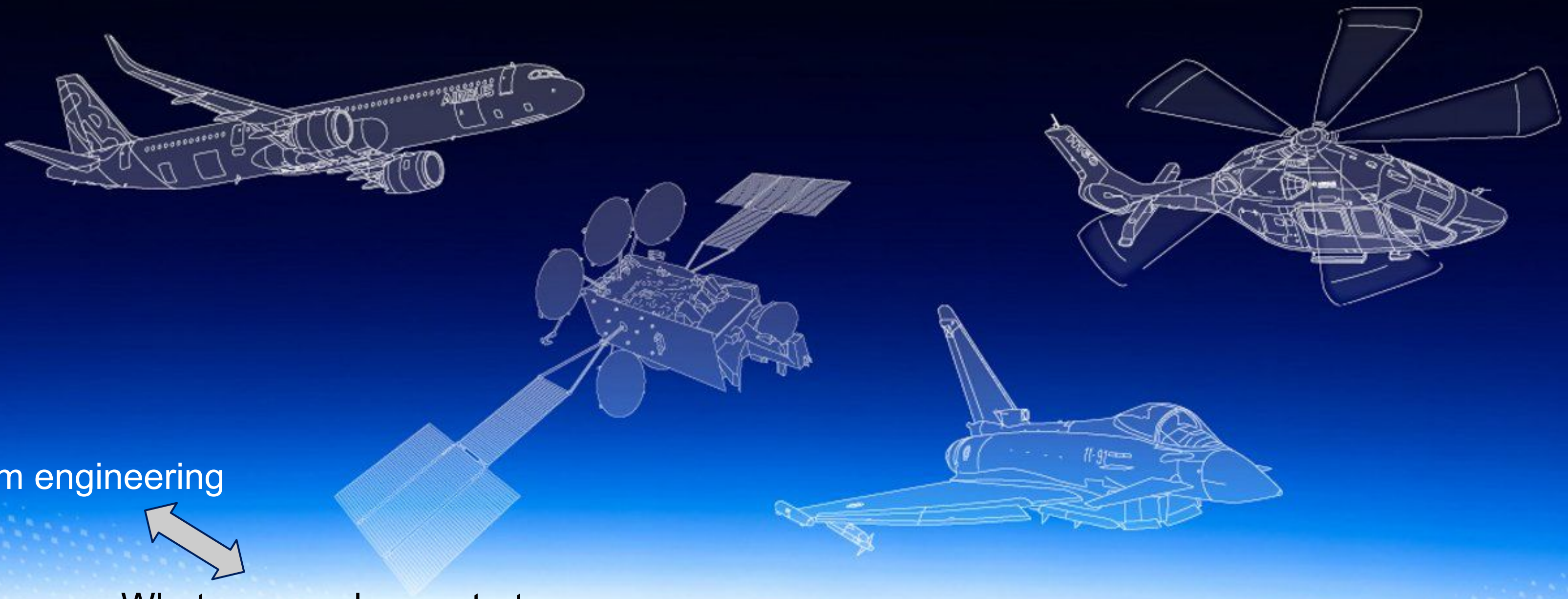
Conclusion (1/2)

	Performance	Levels of Guarantees
Symbolic AI Rules written by humans	+	+++
Data Driven AI Learnt from data	+++	+



AIRBUS

Conclusion (2/2)



Trusted AI methods

Integrate AI in critical system engineering

What we can demonstrate

What we should demonstrate

Ensure Safe and Efficient operations

Academics + Tech companies

Industrials

Regulation authorities