DEEL AI Certification workgroup

23/03/2022

Claire Pagetti

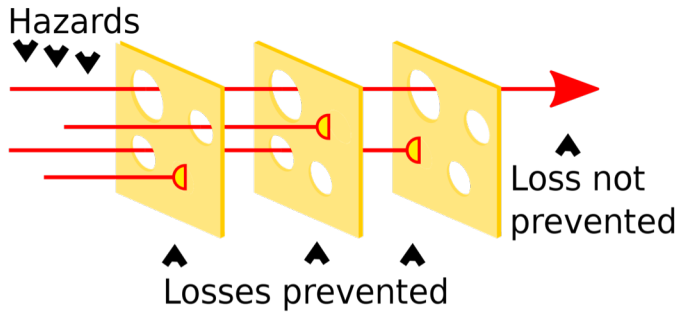
DEEL
DEpendable & Explainable Learning

ANITI
Université Fédérale
Toulouse Midi-Pyrénées

Certification for safety critical systems

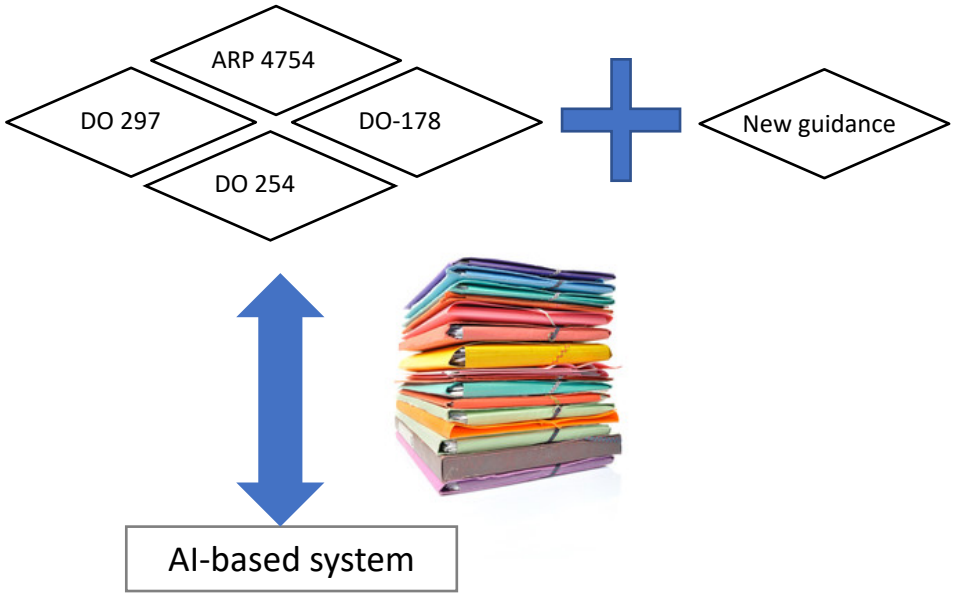
ANITI Domain of application: **transport**

- Safety critical systems
- Safety first



Certification:

- evaluation of an **argumentation**, to convince that a system (i.e., its architecture, its settings, including mitigation means. . .) is compliant with the regulatory requirements
- accepted mean of compliance with the requirements is to rely on **mature standards**



Problem

- Existing certification standards do not address ML-based products (notably the learning phase)
- Theoretical results on ML missing, not mature, included in rich literature always evolving …

Outline

- **Group presentation**
- White paper
- Bottom up activities relying on use cases

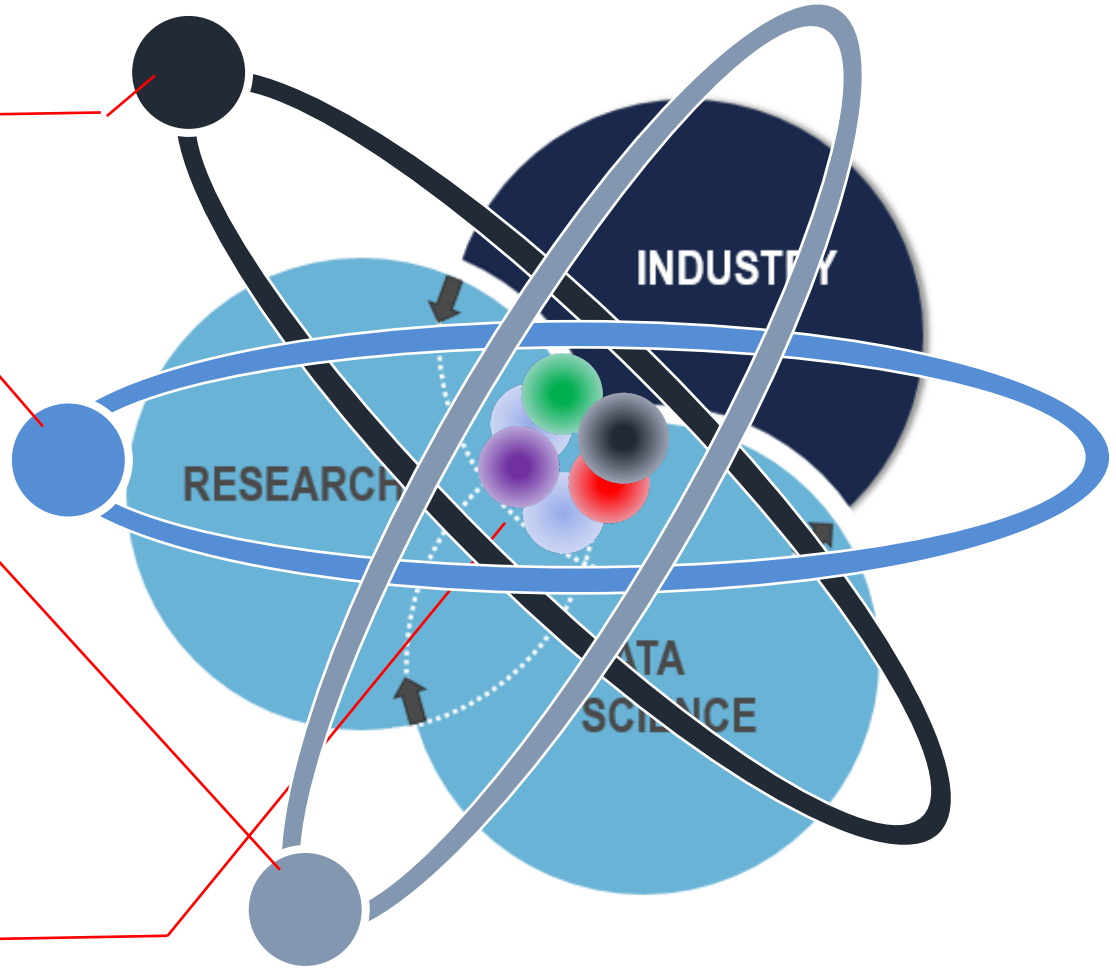
Collaborative environment

- **Kick Off: April 2019**

Expert of safety critical embedded systems

Certification experts in aeronautics, automotive, railway, energy domains

AI experts



DEEL core team: AI experts from industry, researchers, data scientists

Objectives

- **Share knowledge** on certification and ML
  - 48 acculturation sessions (on certification between sectors, between certification and AI experts)
  - Outcomes: skill improvement of all sides
- **Identify the main difficulties** raised by the usage of ML in safety critical systems
  - Working sessions, bibliography
  - Outcomes:
    - white paper and identification of important challenges
- **Feed the Core team** with relevant scientific challenges
  - Tackled challenges: robustness, conformal prediction, explainability
  - Outcomes:
    - Organization of ML Certified Systems workshop, publications
- **Provide elements of confidence for certification**
  - Practical application on use cases via sprint
  - Outcomes :
    - Input to certification working group (EUROCAE WG 114 / SAE G34, SOTIF)
    - Transfer to industries
    - Publications

Outline

- Group presentation
- **White paper**
- Bottom up activities

White Paper

Machine Learning in Certified Systems

DEEL Certification Workgroup
IRT Saint Exupéry
March 2021
Ref - S079L03T00-005

Published on March 2021

https://arxiv.org/abs/2103.10529
https://hal.archives-ouvertes.fr/hal-03176080v1

Approach

Problem

Existing certification standards do not address ML-based products (notably the learning phase)
What are the certification objectives for those products?

Idea

- Identify **properties** that, have a **positive impact on the capability to certify**
- Identify the **main challenges** for the demonstration of **compliance** with the certification objectives

Solution : High Level Properties

- Interpretability / Explainability
- Data Quality
- Robustness
- Repeatability
- Maintainability
- Auditability
- Specifiability
- Verifiability
- Provability
- Resilience
- Monitorability

Main challenges identified

#1 Probabilistic assessment
#2 Resilience
#3 Specifiability
#4 Data Quality and Representativeness
#5 Explainability
#6 Robustness
#7 Verifiability

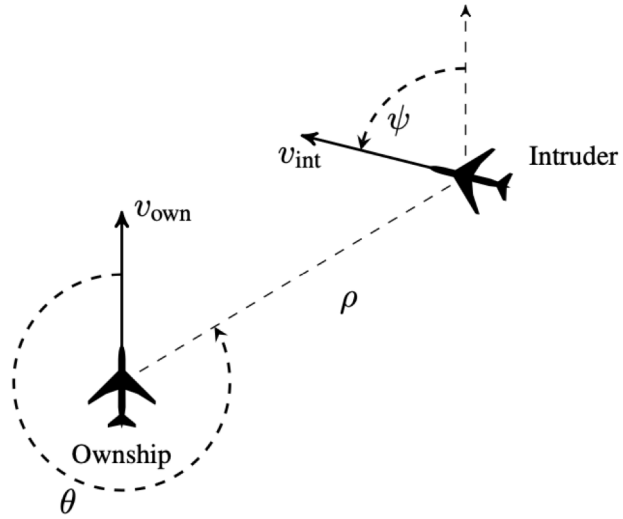
Outline

- Group presentation
- White paper
- **Bottom up activities relying on use cases**

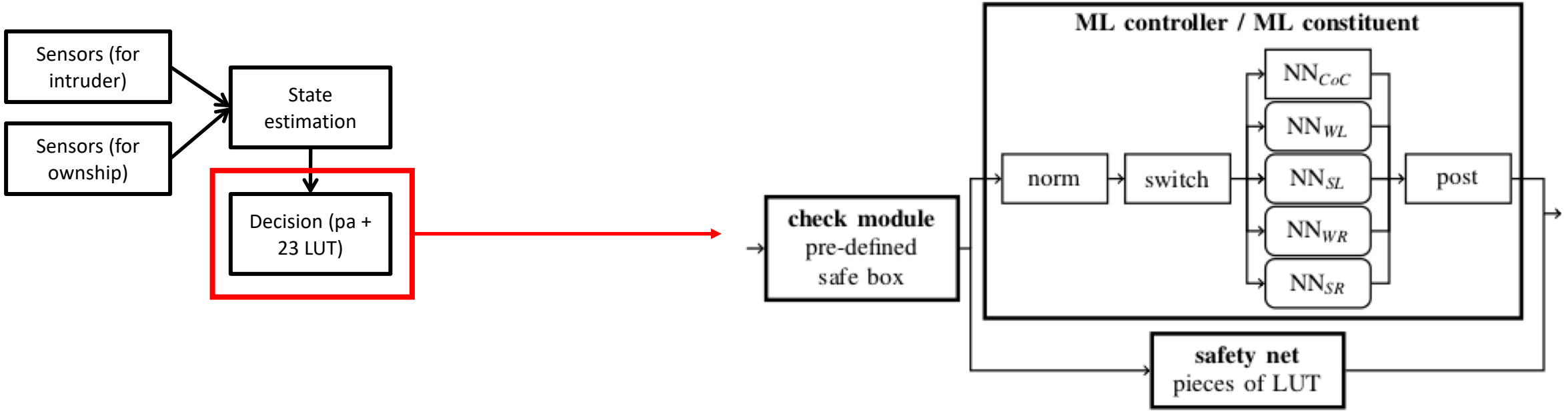
ACAS Xu use case

- Use case
  - avoidance system initially based on look-up tables (LUT)
  - Replacing the LUT by neural networks (cf Reluplex paper)

- Objective:
  - Could we propose a certification strategy for the ML-based system?
  - Interest of the use case: simplest situation (surrogate model with golden truth)

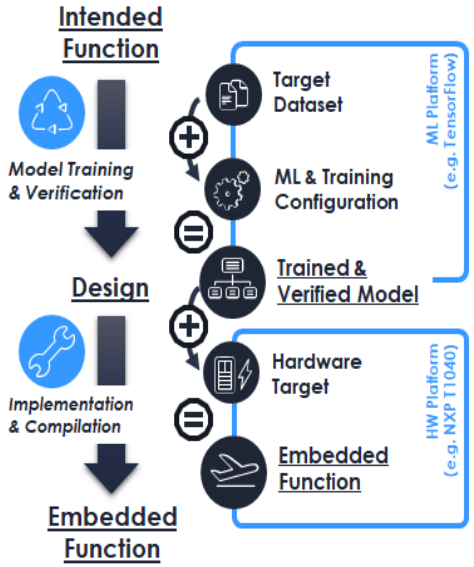


- Results
  1. Definition of a safe architecture (named hybrid architecture)



ACAS Xu use case

- Results
  2. Use of complementary methods (formal methods, simulation …)
     - Formal methods
       - Abstract interpretation (ERAN, Crown, reluval…)
       - Exact solvers (Reluplex/Marabou, Planet …)
     - Development of an ACAS Xu simulator
     - Optimization experiments (pruning, quantization …)

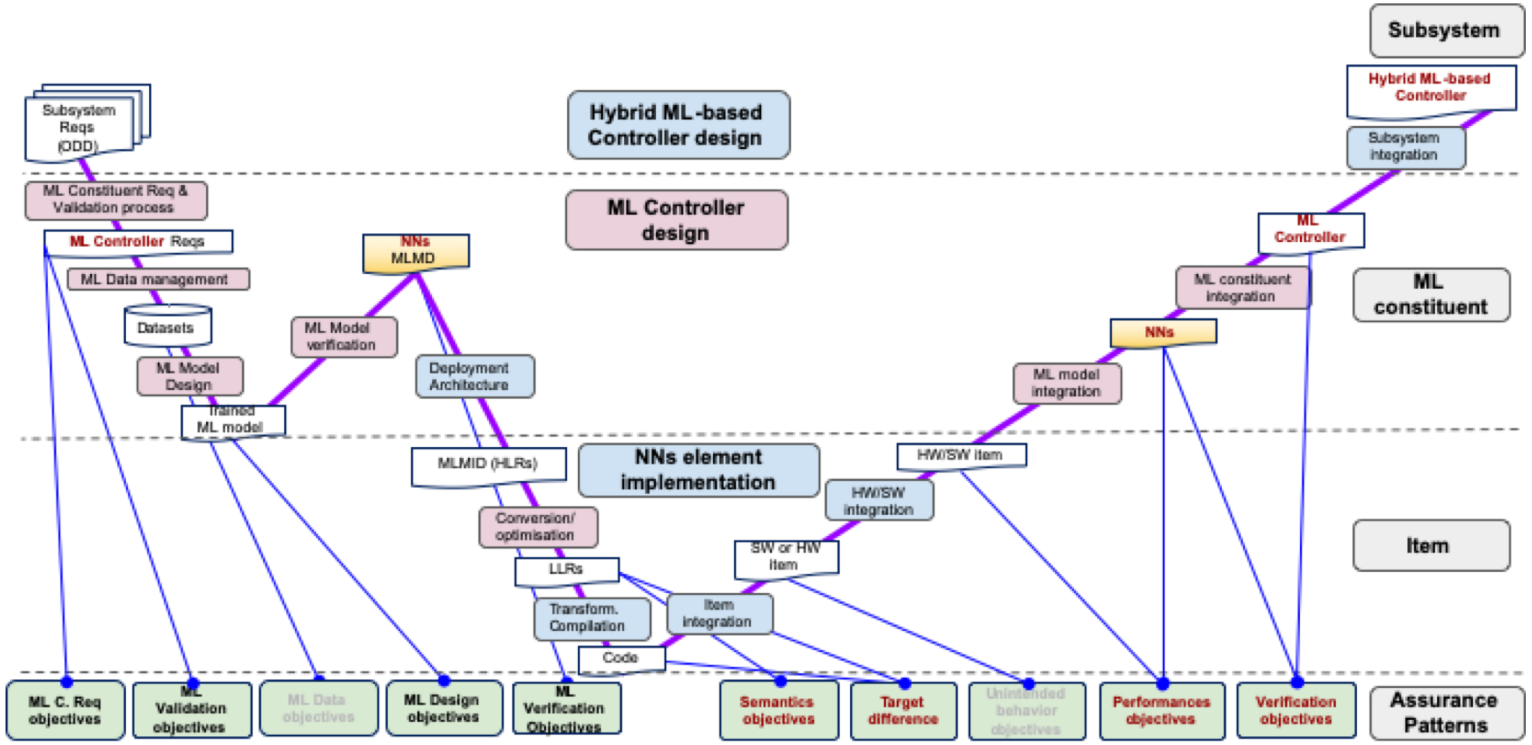


DESIGN
- **Pruning**
- **Quantization**
- Architecture optimization
- Weights binarization

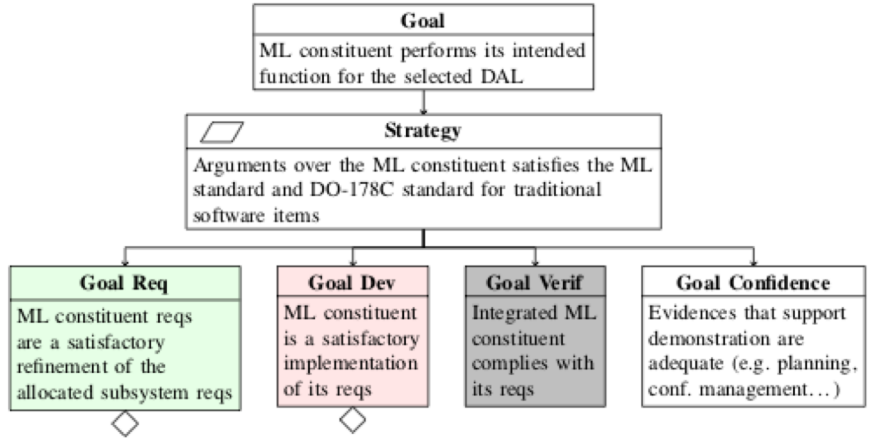
IMPLEMENTATION
- Merging layers
- Winograd
- Fast algorithms (FFT for CONV)
- **Sparsity**
- Parallelization

ACAS Xu use case

- Results
  3. Definition of process and its associated assurance cases



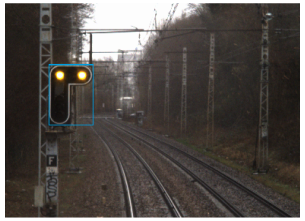
W development process



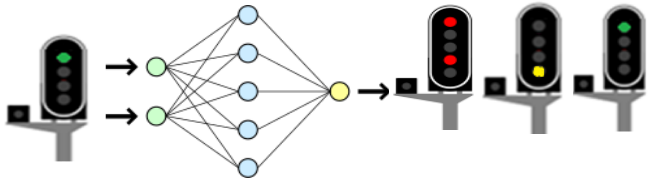
Part of the assurance case

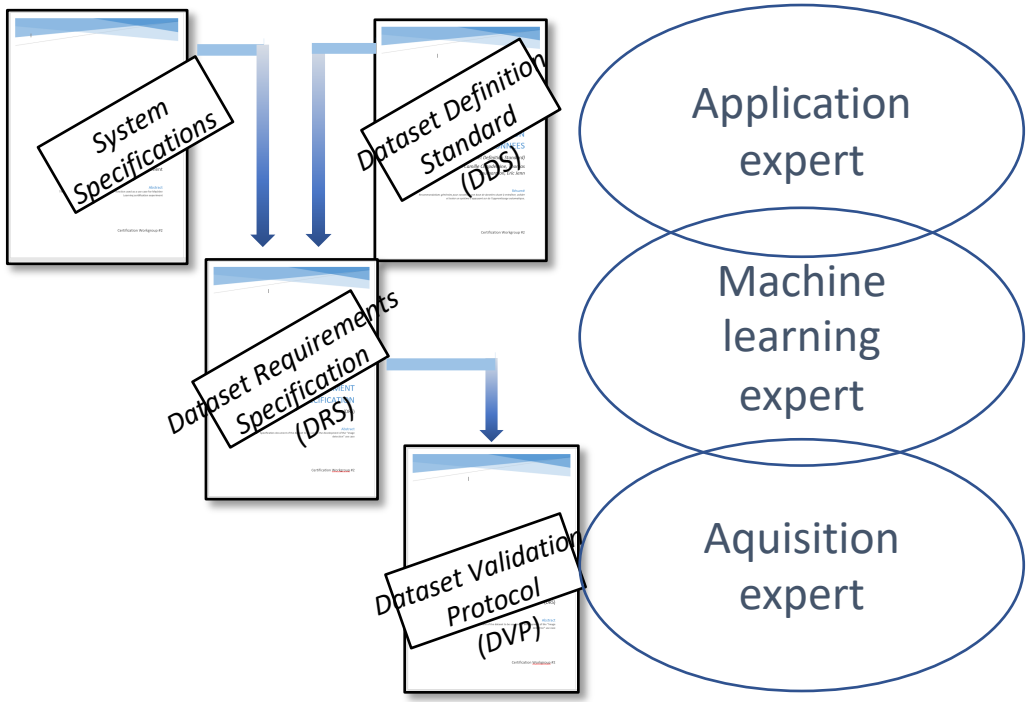
Computer vision use case

- Use case
  - recognition of railway traffic signals with computer vision
- Objective:
  - Could we propose a certification strategy for the ML-based system?
- Results
  1. Dataset quality for machine learning



?

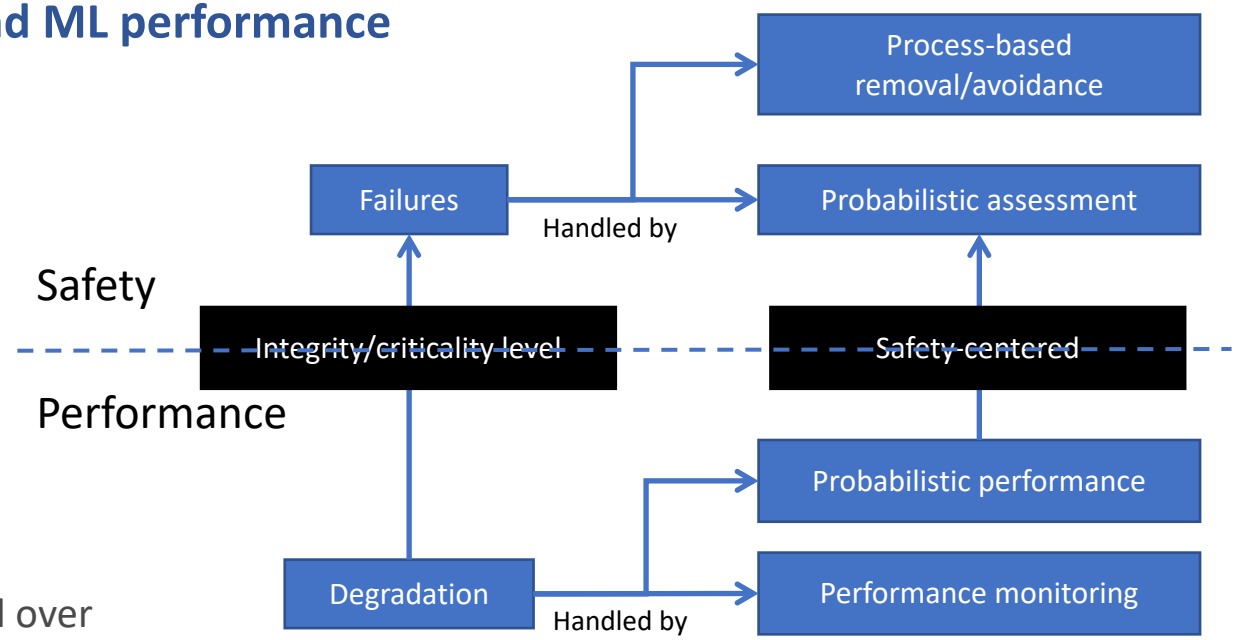




Computer vision use case

- **Results**
  - **2. Link between safety probabilistic assessment and ML performance**

- **Finding: many sources of uncertainty**
  - Dataset uncertainties (e.g. Inter-dataset distribution shift), Training uncertainties (e.g loss function) …
  - Process based development insufficient
  - Idea: computation of statistical guarantees mixed with safety probabilistic approach



- **Contribution:**
  - Formulation 1: on average over all possible training sets and over all future samples
    - Standard formulation (from conformal prediction)
  - Formulation 2: on average for a particular training set and over all future samples
    - Ducoffe et al. 2019
    - Refinement of formulation 1, more samples needed to be statistically relevant
  - Formulation 3: on average for a particular training set and for a given sample
    - Ideal one
    - Knowledge of the input distribution required

Publications

- Christophe Gabreau, Adrien Gauffriau, Florence de Grancey, Jean-Brice Genestet and Claire Pagetti. Assurance Case: A means to support certification of safety-related systems using ML. European Congress on Embedded Real Time Systems (ERTS 2022)
- Hugues Bonnin, Eric Jenn, Lucian Alecu, Thomas Fel, Laurent Gardes, Sébastien Gerchinovitz, Ludovic Ponsolle, Franck Mamalet, Vincent Mussot, Cyril Cappi, Kevin Delmas and Baptiste Lefevre. Can we reconcile safety objectives with machine learning performances? European Congress on Embedded Real Time Systems (ERTS 2022)
- Mathieu Damour, Florence De Grancey, Christophe Gabreau, Adrien Gauffriau, Jean-Brice Ginestet, Alexandre Hervieu, Thomas Huraux, Claire Pagetti, Ludovic Ponsolle, Arthur Clavière. Towards Certification of a Reduced Footprint ACAS-Xu System: A Hybrid ML-Based Solution. International Conference on Computer Safety, Reliability, and Security (SafeComp) 2021
- Eric Jenn, Alexandre Albore, Franck Mamalet, Grégory Flandin, Christophe Gabreau, Hervé Delseny, Adrien Gauffriau, Hugues Bonnin, Lucian Alecu, Jérémy Pirard, Baptiste Lefevre, Jean-Marc Gabriel, Cyril Cappi, Laurent Gardès, Sylvaine Picard, Gilles Dulon, Brice Beltran, Jean-Christophe Bianic, Mathieu Damour, Kevin Delmas, Claire Pagetti. Identifying challenges to the certification of machine learning for safety critical systems. European Congress on Embedded Real Time Systems (ERTS 2020)
- Sylvaine Picard, Camille Chapdelaine, Cyril Cappi, Laurent Gardes, Eric Jenn, Baptiste Lefevre, Thomas Soumarmon. Ensuring dataset quality for machine learning certification. 2020 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW)

DEEL
DEpendable & Explainable Learning

QUESTIONS

IVADO
IRT SAINT EXUPÉRY
CRIAQ
ANITI ARTIFICIAL & NATURAL INTELLIGENCE TOULOUSE INSTITUTE
Institut intelligence et données