

# PANEL 1

## CERTIFICATION OF MACHINE LEARNING FOR SAFETY CRITICAL APPLICATIONS: PROBABLE, PLAUSIBLE OR IMPRACTICAL?



# Panel overview

**Moderator:** Claire Pagetti (ONERA/ENSHEEIT-IRIT/ANITI)

**Panelists:**

**Hugues Bonnin** Senior Expert Continental Digital Service

**Loïc Correnson** Senior Expert CEA

**Christophe Gabreau** Airbus - Eurocae WG114 group leader

**Franck Mamalet** DEEL "certification of AI" mission co-animator

**Guillaume Soudain** EASA – Software Senior Expert

**Agenda:**

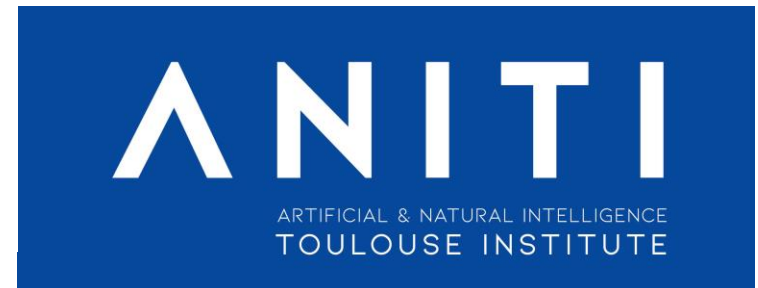
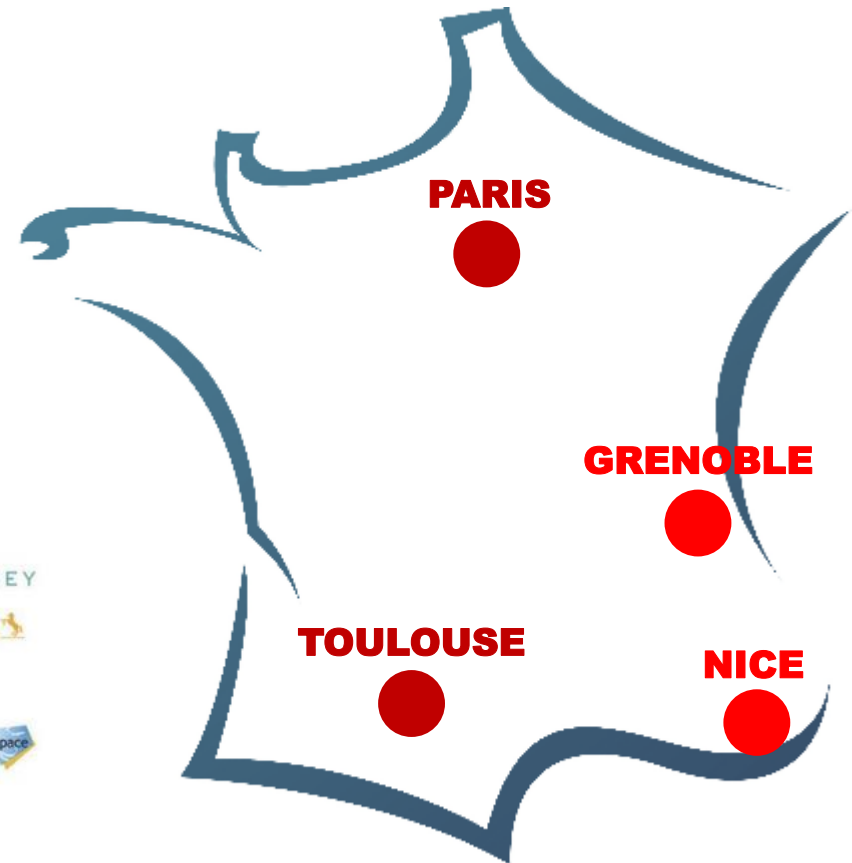
- Brief introduction of the panel and Toulouse ANITI project
- 30 min: short presentation by each panelist
- 30 min: questions from the assistance

# Toulouse eco-system

- ❑ ANITI (Artificial and Natural Intelligence Toulouse Institute)

<https://aniti.univ-toulouse.fr>

- ❑ One of the 3iA: Interdisciplinary Institutes for AI
- ❑ +50 partners



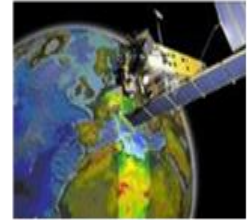
# ANITI Ambition

Enable the use and the long-term development of AI in human critical sectors (transport...) and industry by addressing following key challenges

*Acceptability*

*Fair Data*

*Reliability*

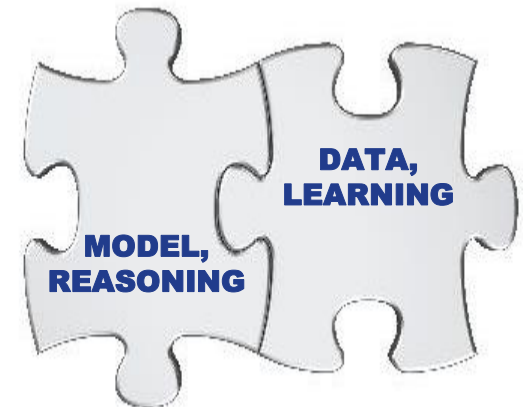


*Adaptability*

*Explainability*

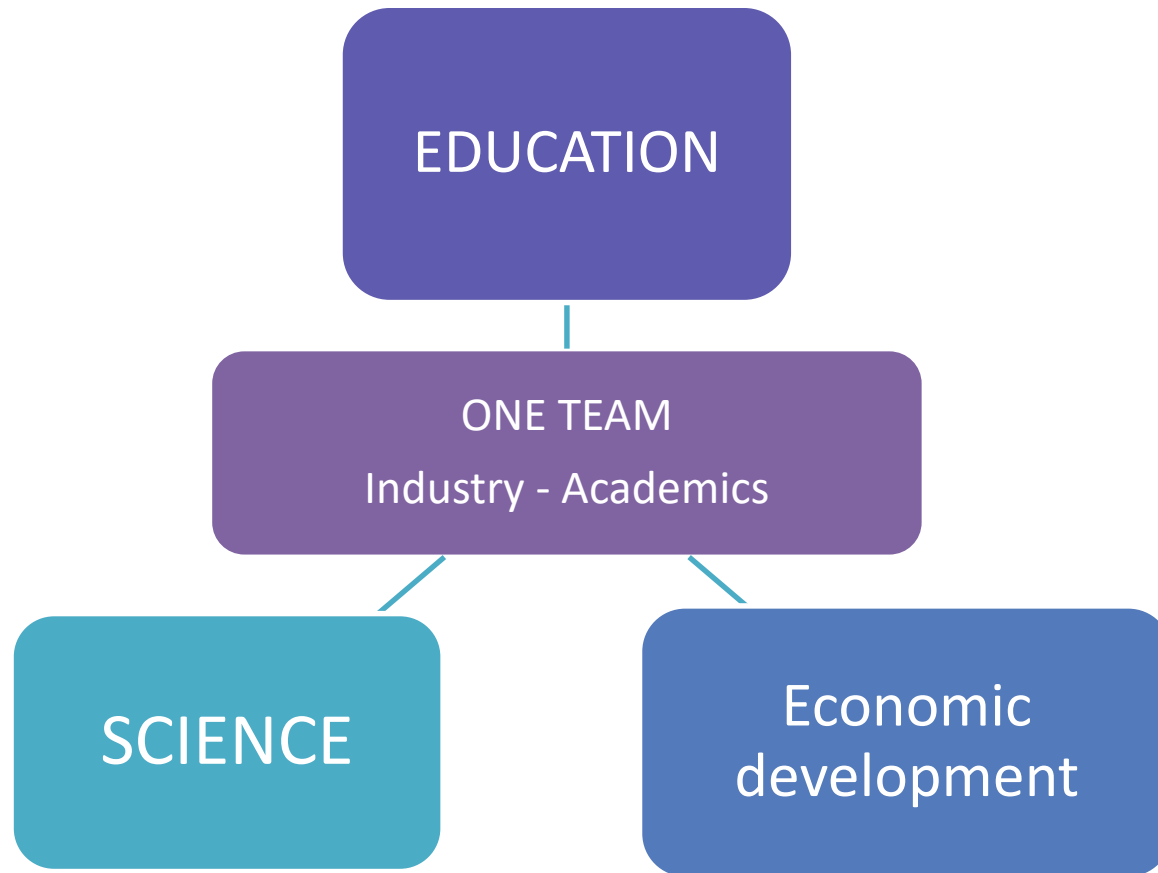
*Scalability*

**Hybrid approach mixing Model-based and Data-based IA is an efficient way to address above challenges**



# How to get there ?

## 3 complementary pillars and a collaborative way-of working



# Panel overview

**Moderator:** Claire Pagetti (ONERA/ENSHEEIT-IRIT/ANITI)

**Panelists:**

**Hugues Bonnin** Senior Expert Continental Digital Service

**Loïc Correnson** Senior Expert CEA

**Christophe Gabreau** Airbus - Eurocae WG114 group leader

**Franck Mamalet** DEEL "certification of AI" mission co-animator

**Guillaume Soudain** EASA – Software Senior Expert

**Agenda:**

- Brief introduction of the panel and Toulouse ANITI project
- 30 min: short presentation by each panelist
- 30 min: questions from the assistance

# Continental Digital Service



A yellow measuring tape with black markings is coiled into a spiral shape on a black background. The tape has numbers in both inches and centimeters. The word "CERTification" is overlaid in blue text.

**CERTification**

A black swan with a red beak is swimming in a pond, surrounded by many white swans. The word "unCERTainty" is overlaid in blue text.

**unCERTainty**



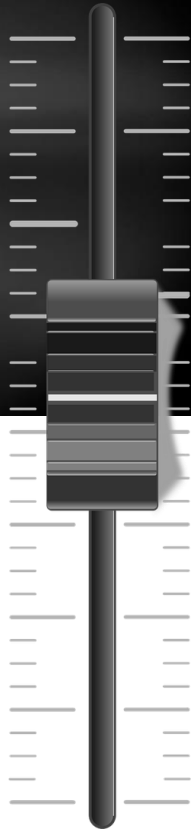
## ❑ Ensuring safety by **surrounding** the ML model:

- Black box approach
- Exploring the unknowns, unknowns unknowns...
- Reducing the unknowns risky zones



## ❑ **building** the model with properties that promote the safety

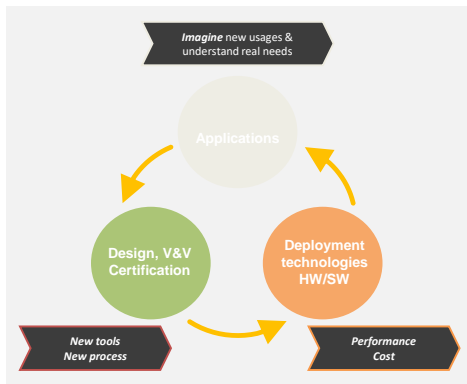
- Explainability (?)
- Level of confidence
- “safe” cost function
- Methodical dataset choice and build



# CEA List

--	--

# AI CERTIFICATION, CEA activities cover the whole scope



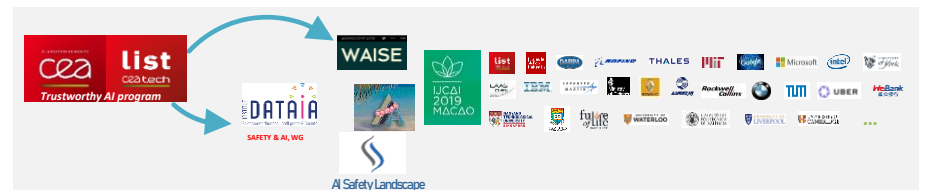
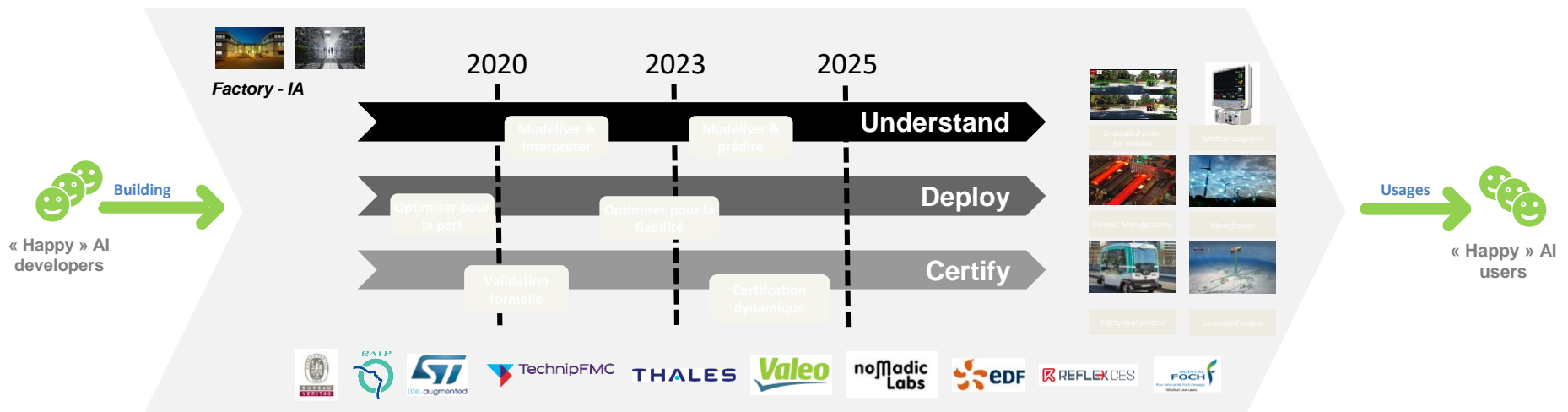
« **TRUST**  
will make the  
difference

(Quality, Safety, Reliability,  
Ethic, Responsibility...)

Trust:

A top level strategy of CEA, validated by CEA's Ministeries and  
leaded by CEA List on the basys of its DNA on digital trust

# A ROADMAP to provide a CHAIN OF TOOLS & TECHNOLOGIES for Reasearch & Development of CERTIFIED AI



# WG-114: Jan 2020 status



Christophe Gabreau (co-chair)  
Christian Thurow (co-chair)



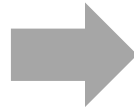
# Standardization Path

THALES  Rolls-Royce  
 BOEING  DASSAULT  
AVIATION  
AIRBUS Honeywell ...



...

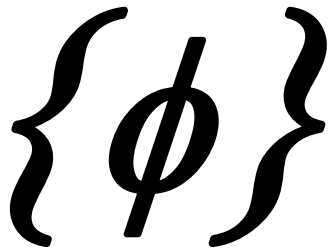
Industry  
Methods  
& Processes



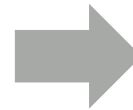
Standard  
Document



Acceptable Means of  
Compliance  
AMC/AC to 25.1309



AI Certification  
Standard  
Document



Acceptable Means of  
Compliance  
AMC/AC to 25.1309

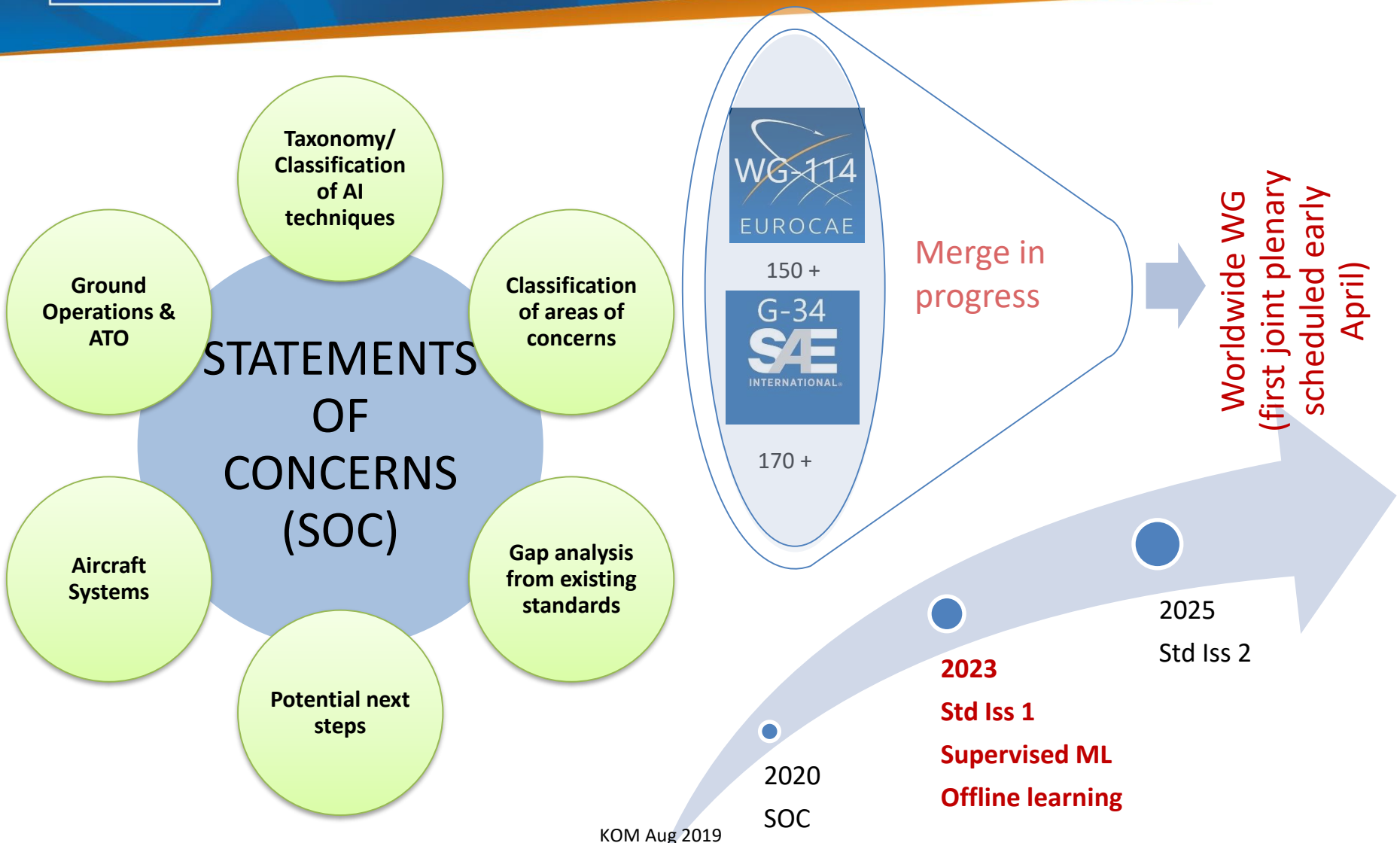


...





# Standardization Path



# DEEL

DEpendable & Explainable Learning

DEpendable & Explainable Learning



# DEEL Project

**DEEL**  
DEpendable & Explainable Learning



The project will provide industrial partners with the artificial intelligence (AI) tools and technological bricks that enable them to **secure and certify** in a short time the development of their critical systems integrating AI functions.

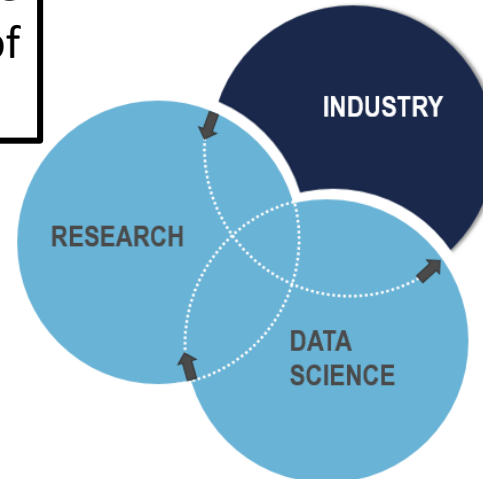
Main topics:

- Explainability,
- Robustness,
- Fairness,
- Certificability,
- Privacy

Toulouse



Montréal



26  
M€

5  
ans

24  
partners

60  
personnes

# DEEL project

**DEEL**  
DEpendable & EXplainable Learning

## Fairness challenge objectives:

- removing bias from training (training data, unfair decisions)
- effects of the learning sample in the ML process use it to improve training datasets
- collaborative training with separate and secret datasets

M. Serrurier, Jean-Michel Loubes, et E. Pauwels, « Fairness with Wasserstein Adversarial Networks », 2019.

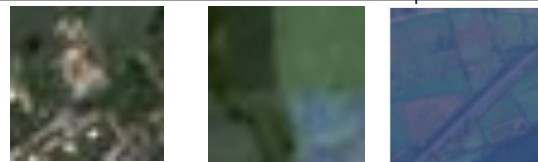
Risser, Vincenot, Couellan, Loubes, 2019 <https://arxiv.org/abs/1908.05783>

## Explicability challenge objectives:

- User concepts for explainability toolbox (industrial and certification level)
- Explainability of black box models
- Stability of models and interpretability
- Metrics

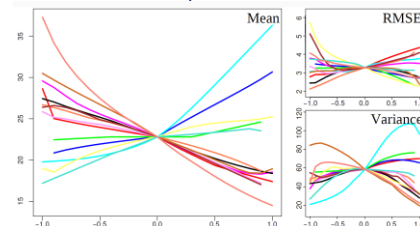
## Deliverables:

- methodology, tools, metrics to measure bias
- algorithms to correct bias in models
- benchmark with industrial applications,



## Deliverables:

- Methodology, tools, metrics for explanation
- Algorithms to explain black box models
- Tutorials on industrial use cases



<https://github.com/XAI-ANITI/ethik>

« Certification of AI » mission presented yesterday (paper We.1.B.3)



[www.mobilit.ai](http://www.mobilit.ai)





# THANKS!





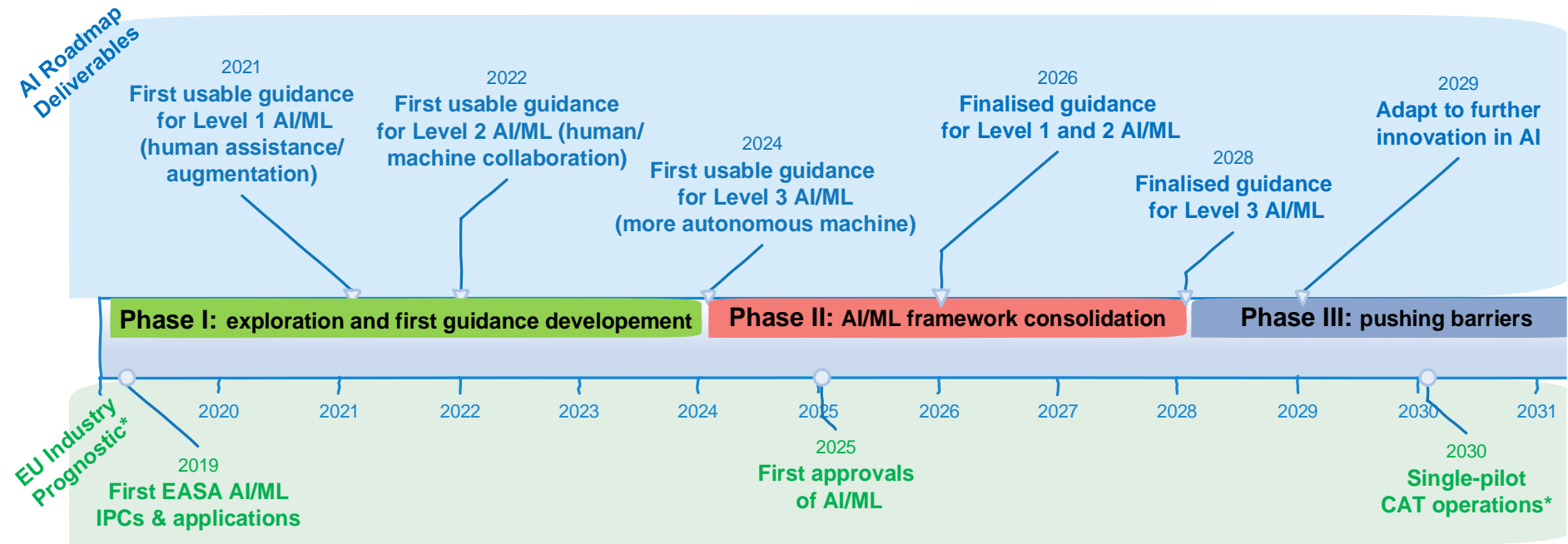
# EASA Artificial Intelligence Roadmap

**ERTS - 30.01.2020**

**Guillaume Soudain – EASA Software Senior Expert**



# AI Roadmap timeline



\*For Large Aircrafts, based on roadmaps from major players

<https://www.easa.europa.eu/document-library/general-publications/european-plan-aviation-safety-2020-2024>

# AI Roadmap building blocks

