DEEL

DEpendable & Explainable Learning

Identifying Challenges to the Certification of Machine Learning for Safety Critical Systems

ERTS 2020

DEEL ML certification workgroup (see next page)
[Speaker: Eric Jenn, eric.jennn@irt-saintexupery.com]

IVADO IRT Saint Exupery CRIAQ ANITI Artificial & Natural Intelligence Toulouse Institute

Composition of the workgroup

Alexandre Albore, Lucian Alecu, Brice Beltran, Hugues Bonnin, Jean-Christophe Bianic, Cyril Cappi, Mathieu Damour, Kevin Delmas, Hervé Delseny, Gilles Dulon, **Grégory Flandin**, Christophe Gabreau, Jean-Marc Gabriel, Laurent Gardès, Adrien Gauffriau, Eric Jenn, Baptiste Lefevre, **Franck Mamalet**, **Claire Pagetti**, Sylvaine Picard, Jérémy Pirard

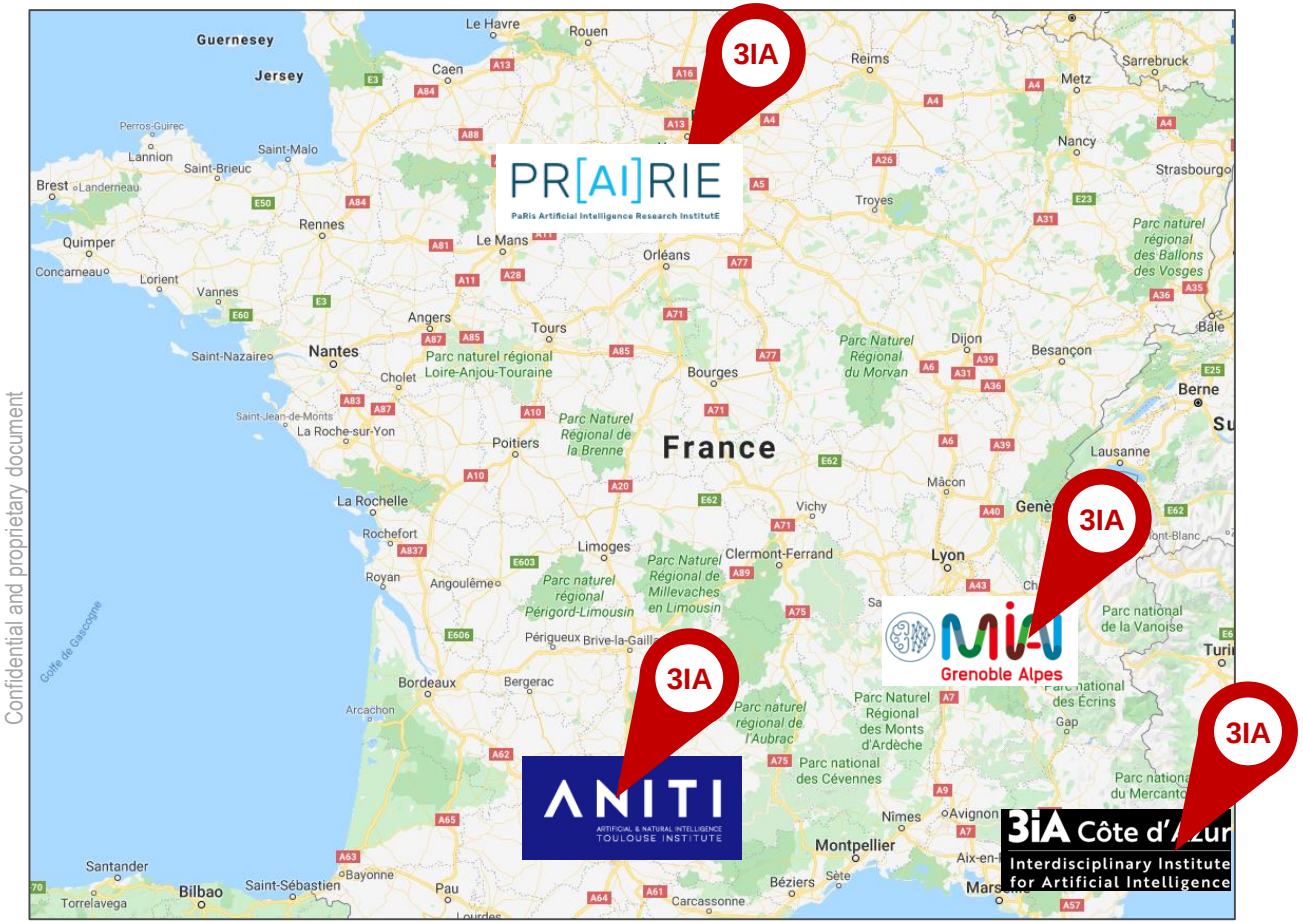
(see paper for affiliations)

- IRT Saint-Exupéry
- AIRBUS SAS
- Continental
- Thales
- Renault Software Labs
- SNCF
- SAFRAN
- DGA
- SCALIAN
- ONERA

OVERVIEW

- The context, the workgroup,…
- About certification…
- New paradigm, new practices,…
- What we've done, how we did it…
- Some challenges…
- Future work

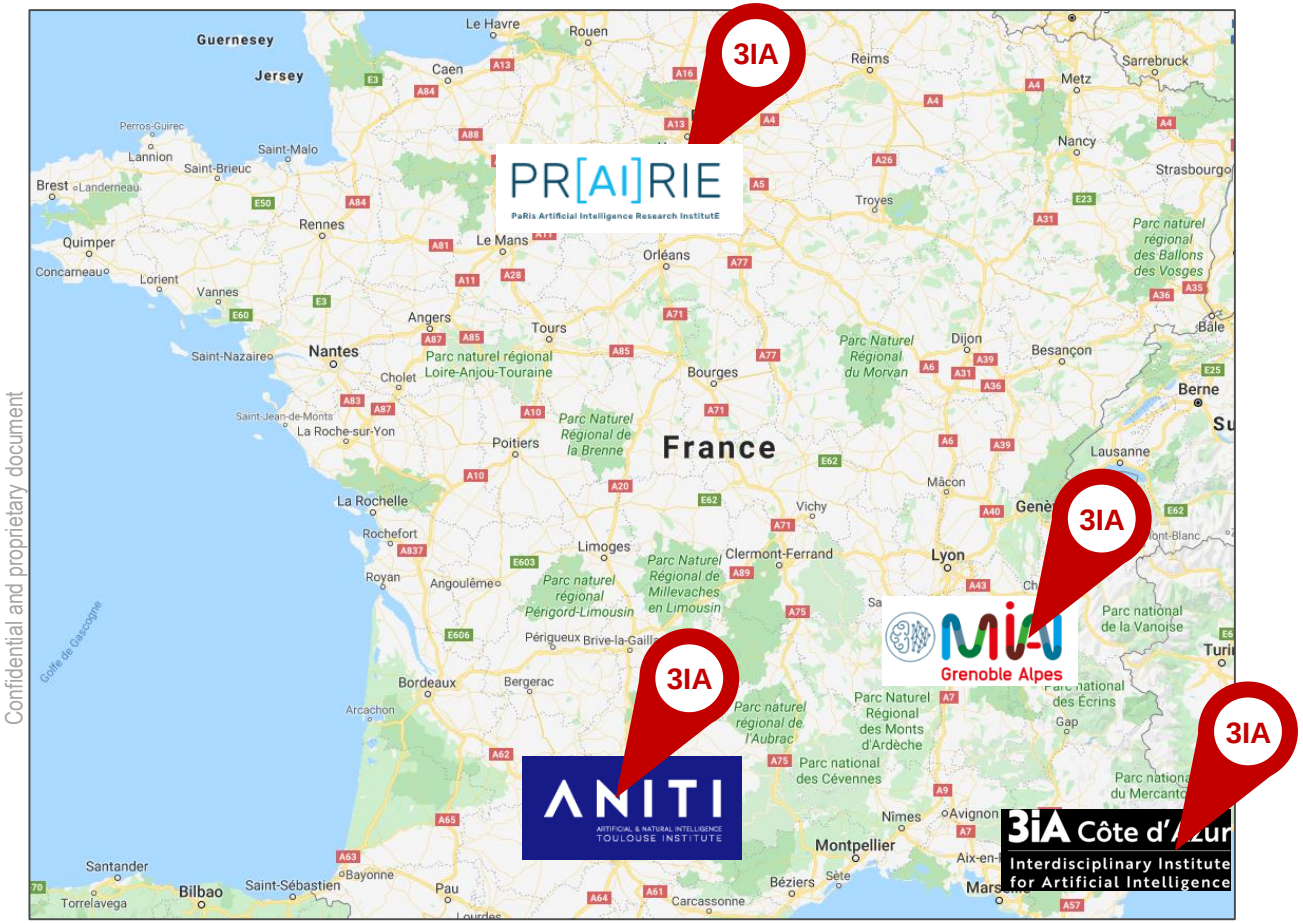
The context



4 Interdisciplinarity Institutes for Artificial Intelligence (3IA):

- Paris
- Grenoble
- Nice
- Toulouse

The context

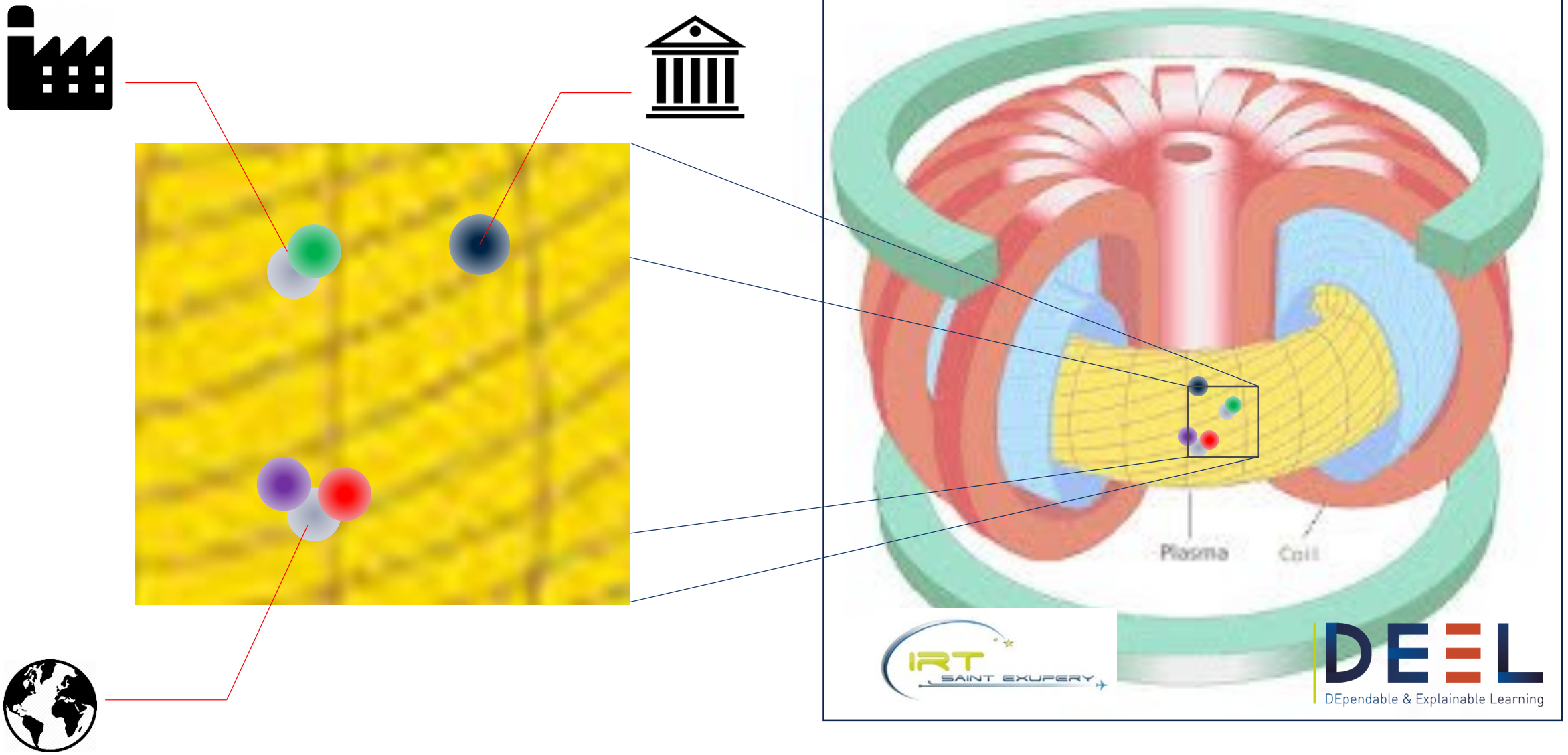


4 Interdisciplinarity Institutes for Artificial Intelligence (3IA):

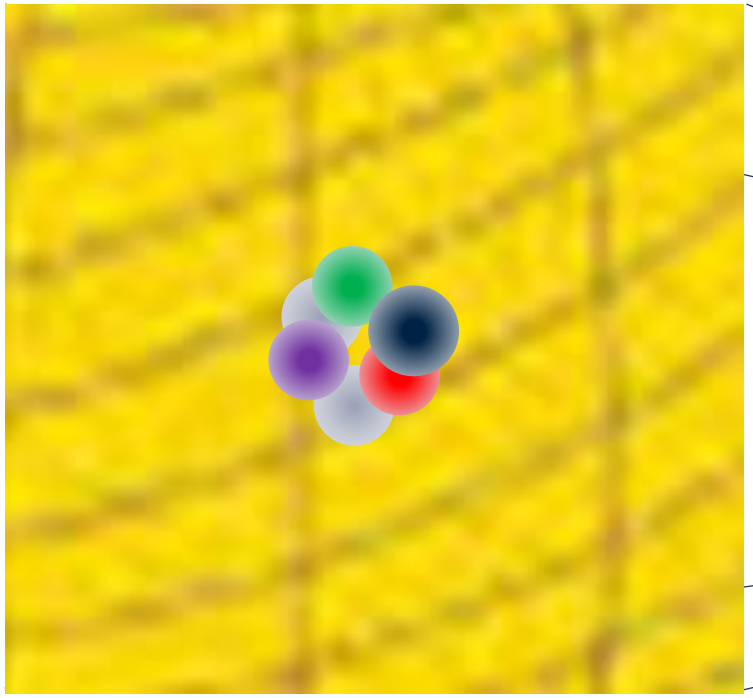
- Paris
- Grenoble
- Nice
- Toulouse

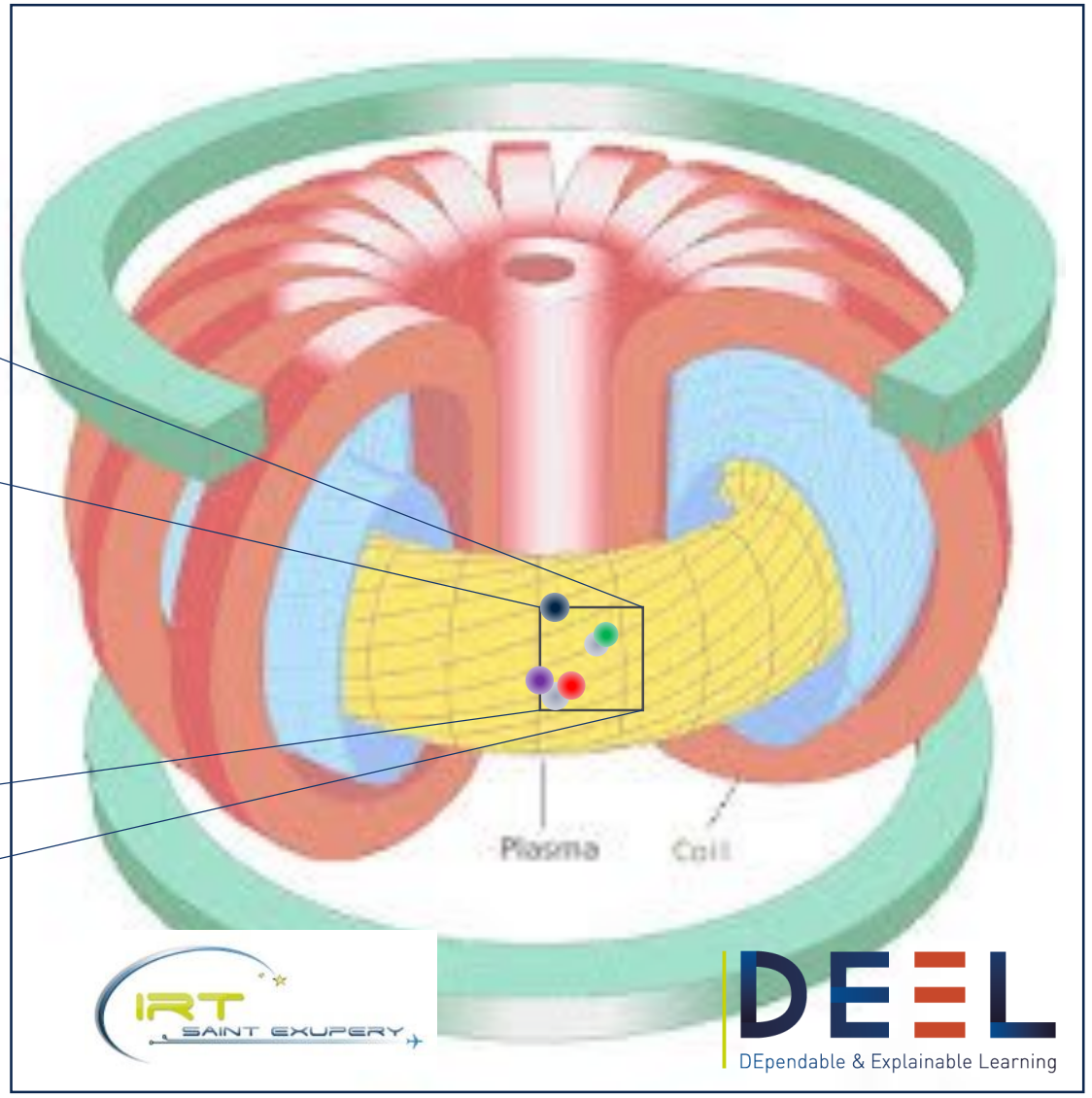
The context

The context

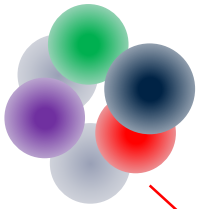


The context



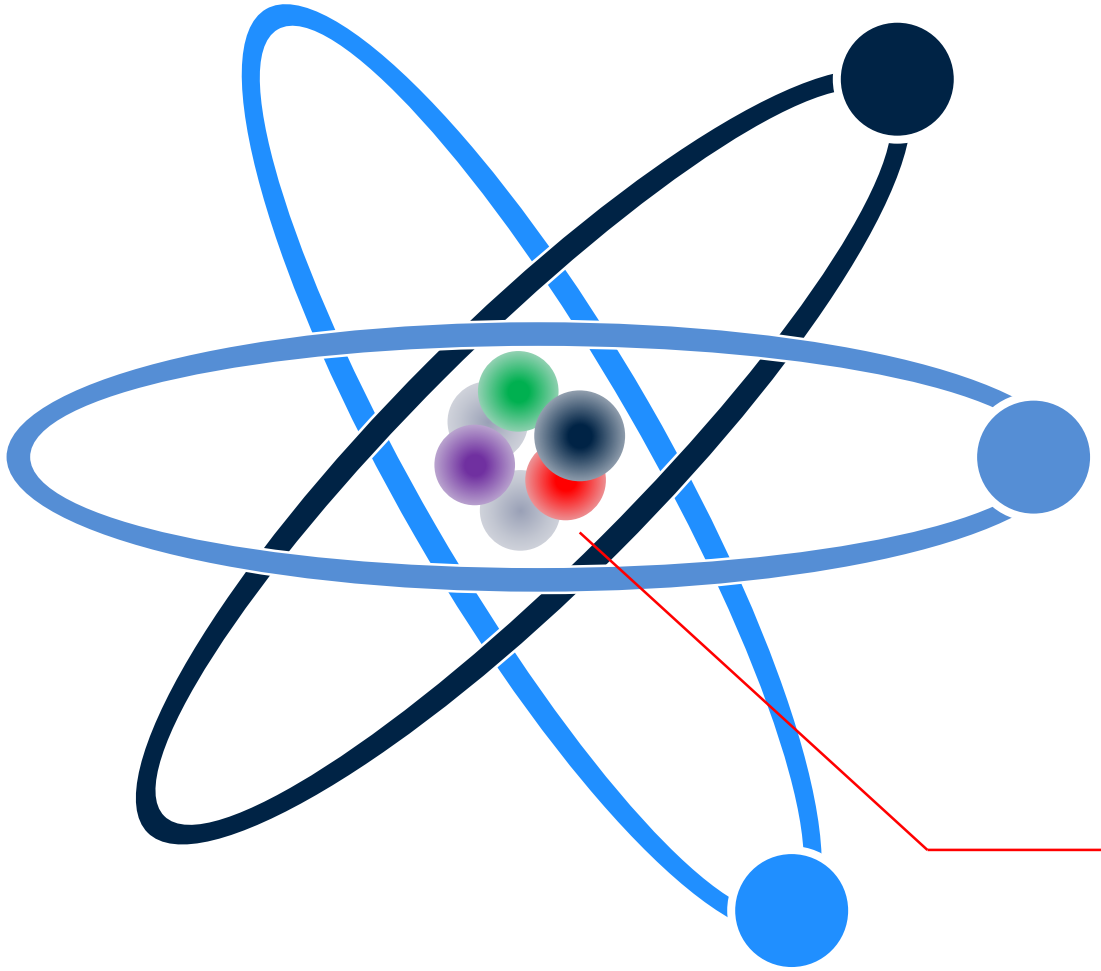


The context



DEEL core team: AI experts from industry, researchers, young engineers

The context

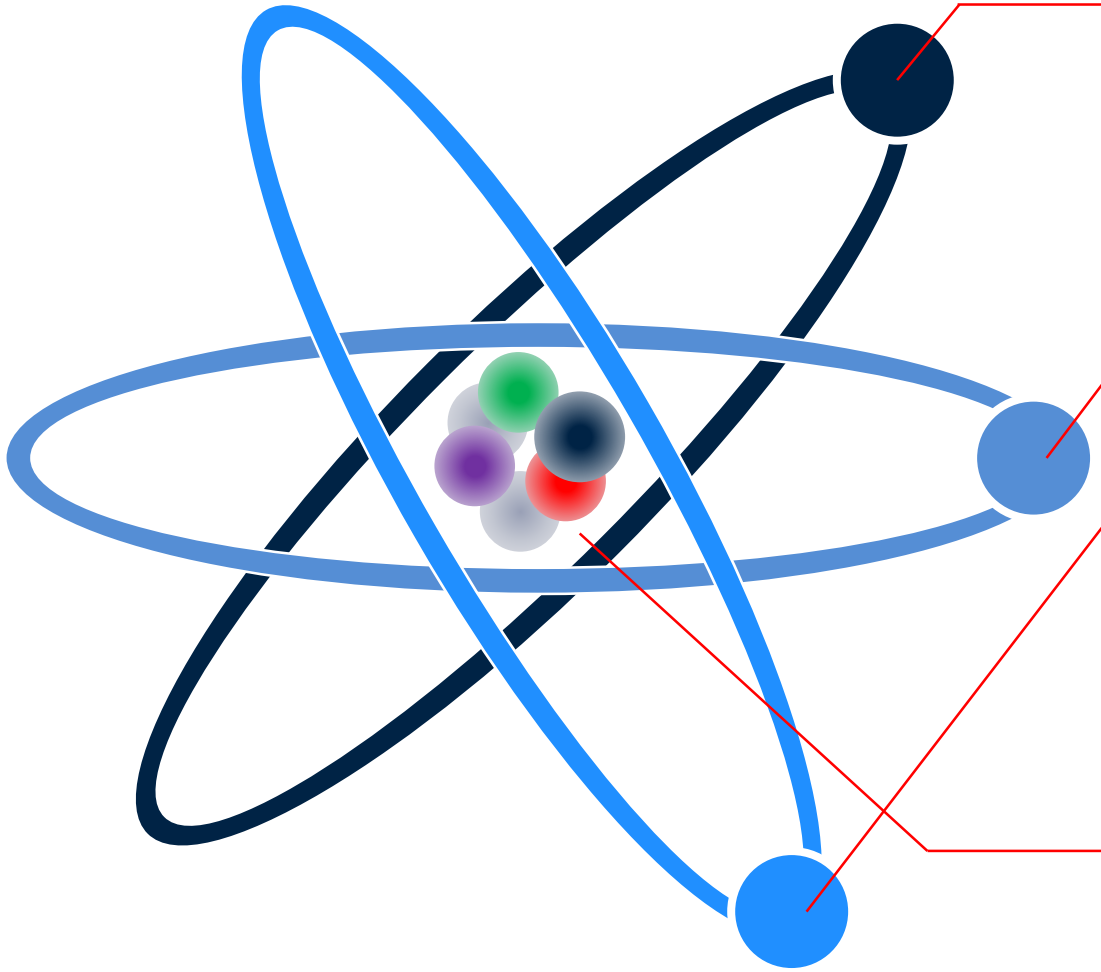


AI CERTIFICATION WORKGROUP

Airbus
Apsys
Continental
DGA
ONERA
Renault
Safran
Scalian:
SNCF
Thales AVS

DEEL CORE TEAM

The context



Expert of safety critical embedded systems

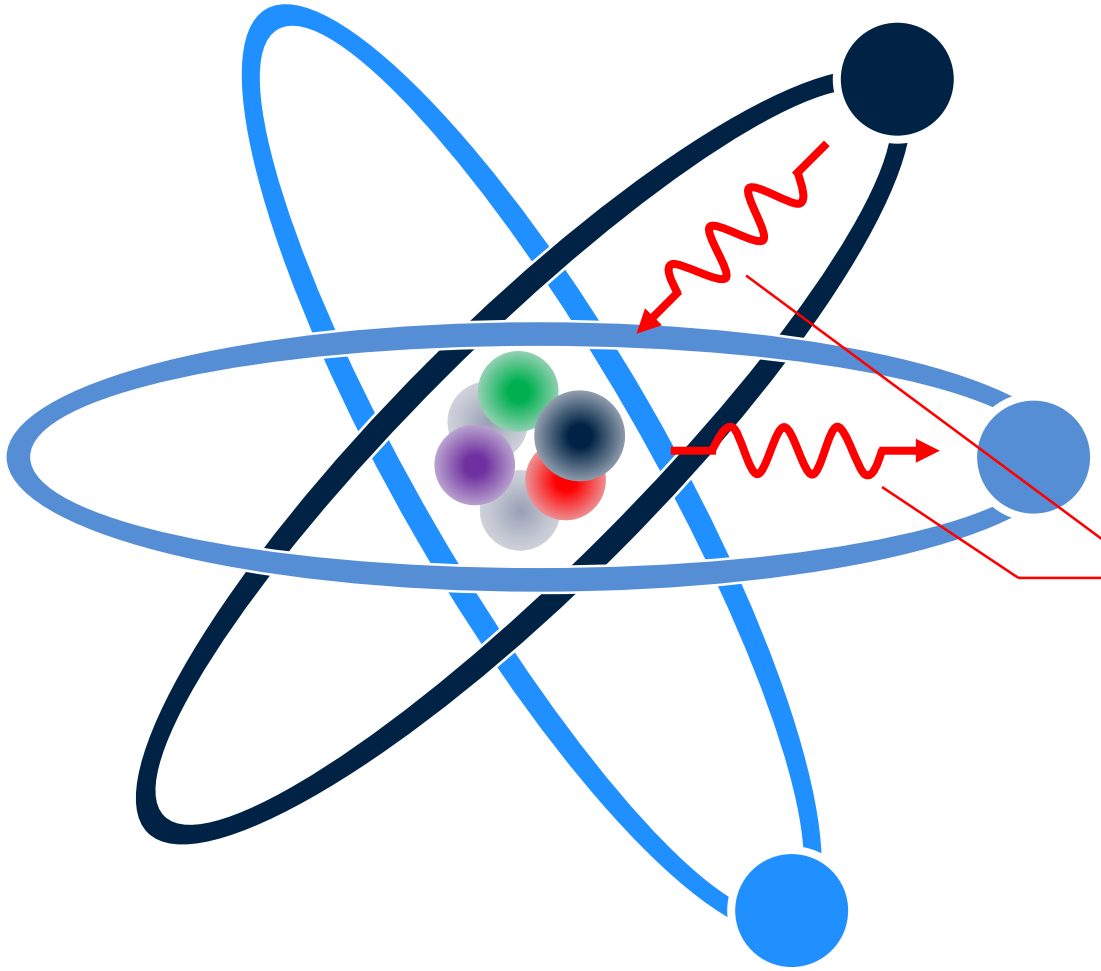
Certification experts in aeronautics, automotive, railway domains

AI experts

Airbus
Apsys
Continental
DGA
ONERA
Renault
Safran
Scalian:
SNCF
Thales AVS

DEEL CORE TEAM

The context



Expert of safety critical embedded systems

Certification experts in aeronautics, automotive, railway domains

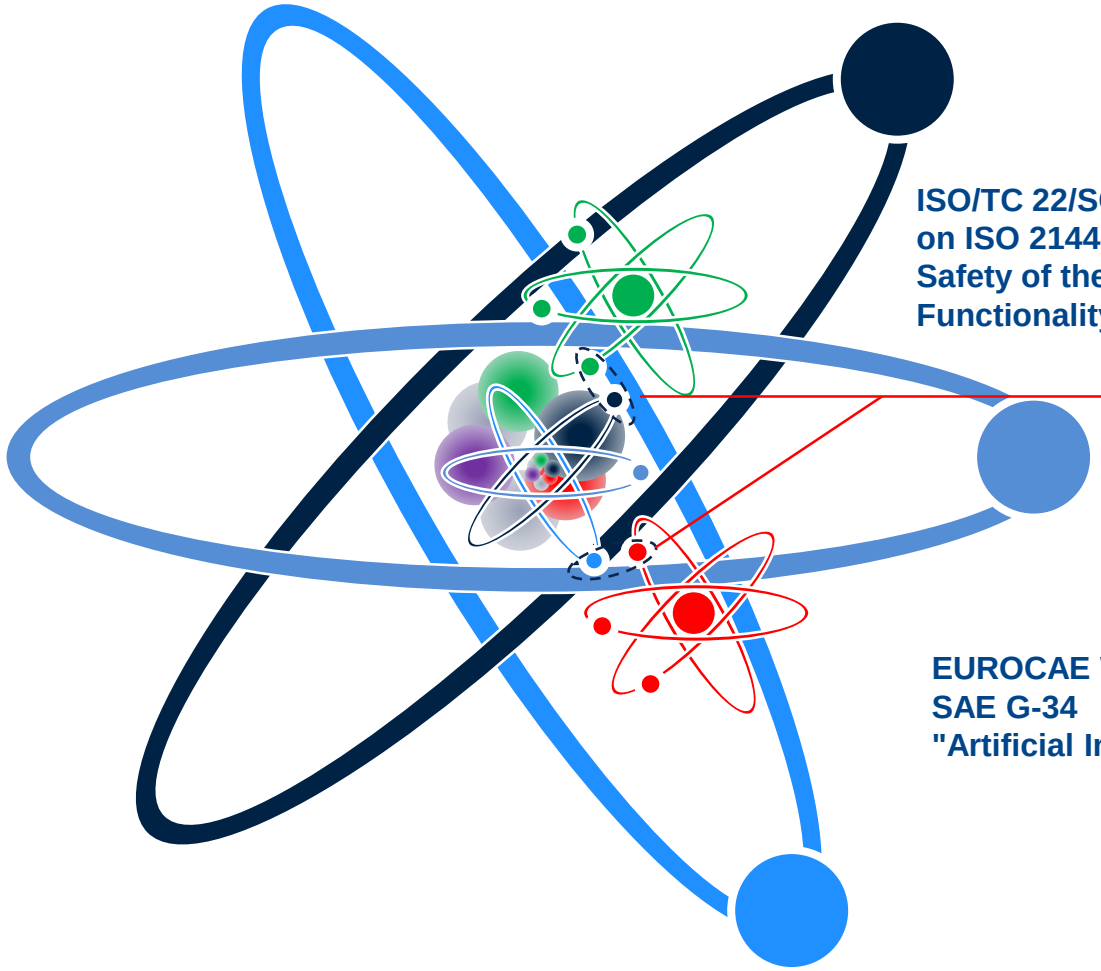
AI experts

Airbus
Apsys
Continental
DGA
ONERA
Renault
Safran
Scalian:
SNCF
Thales AVS

Acculturation
Definition of challenges

DEEL CORE TEAM

The context



ISO/TC 22/SC 32/WG8 working on ISO 21448 Road Vehicles — Safety of the Intended Functionality, aka SOTIF)

EUROCAE WG-114
SAE G-34
"Artificial Intelligence"

AI CERTIFICATION WORKGROUP

Airbus
Apsys
Continental
DGA
ONERA
Renault
Safran
Scalian:
SNCF
Thales AVS

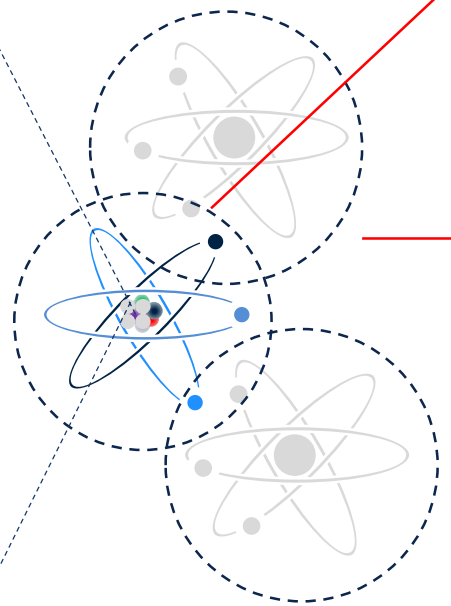
Co-membership

Acculturation
Definition of challenges

DEEL CORE TEAM

The context

White paper (Q1 2020)



ISO/TC 22/SC 32/WG8 working on ISO 21448 Road Vehicles — Safety of the Intended Functionality, aka SOTIF)

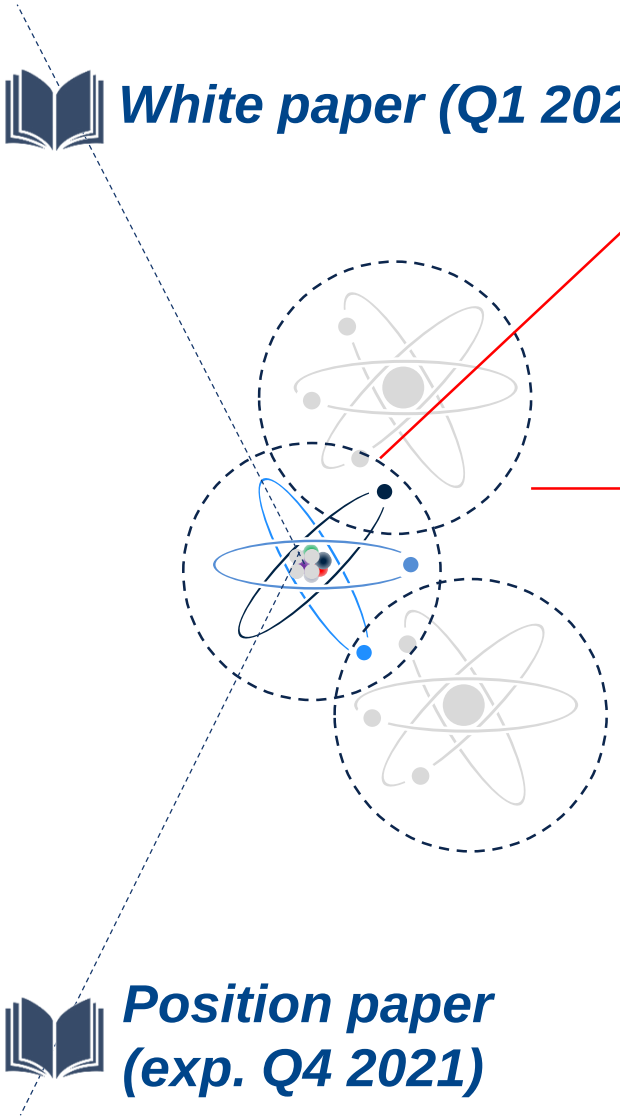
EUROCAE WG-114
SAE G-34
"Artificial Intelligence"

Identify issues to build critical systems embedding ML components
Define research axes to address these issues

Formalisation into processes, recommandations, etc.

Position paper (exp. Q4 2021)

The context



White paper (Q1 2020)

ISO/TC 22/SC 32/WG8 working on ISO 21448 Road Vehicles — Safety of the Intended Functionality, aka SOTIF)

EUROCAE WG-114
SAE G-34
"Artificial Intelligence"

Position paper (exp. Q4 2021)

Identify issues to build critical systems embedding ML components
Define research axes to address these issues

Formalisation into processes, recommandations, etc.

ARPs, DO178, DO254,…
ISO 26262

❑ An (informal) view of the “certification”

CERTIFICATION: The legal recognition that a product, service, organization or person complies with the applicable requirements. Such certification comprises the activity of technically checking the product, service, organization or person, and the formal recognition of compliance with the applicable requirements by issue of a certificate, license, approval or other document as required by national laws and procedures.

- Ensure that a system **does exactly what is it intended to do**...
- Propose a set of **shared, consensual** recommendations in a given domain

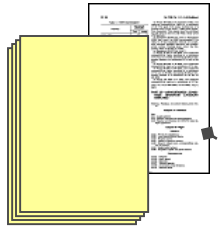
ARPs, DO178, DO254,…
ISO 26262

Certification, aeronautics

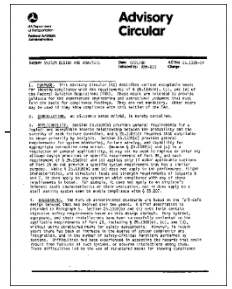
Chicago Convention

Instantiated in US

Code of Federal Regulations

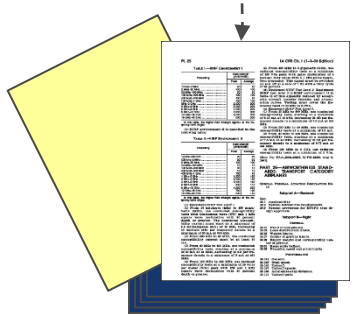


Part 14: Aeronautics and Space

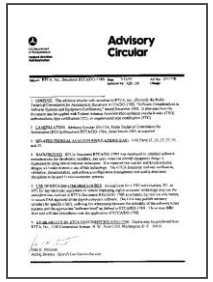


AC 25.1309
"Acceptable Means of Compliance to 25.1309

FAR 25.1309
Part 25 : Subpart F - Equipments



Part 25: Airworthiness standards:
Transport category airplanes



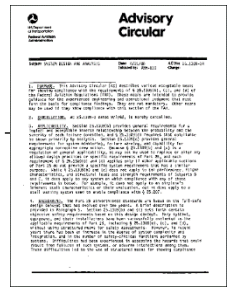
AC 20-115D

ARPs, DO178, DO254,...
ISO 26262

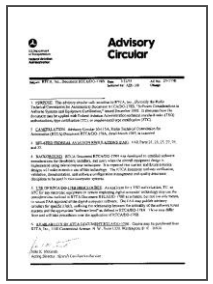
Certification, aeronautics

Chicago Convention

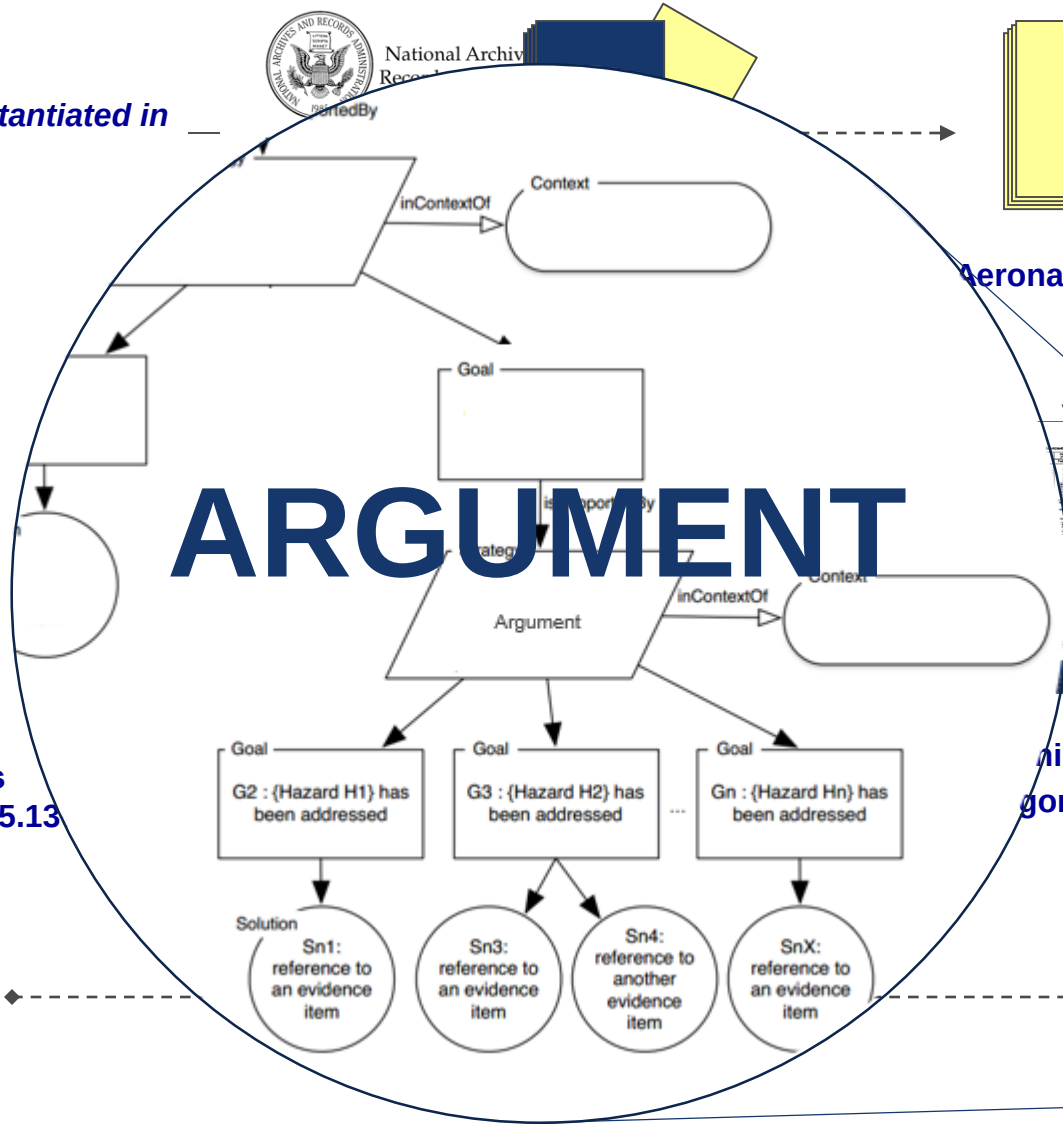
Instantiated in US

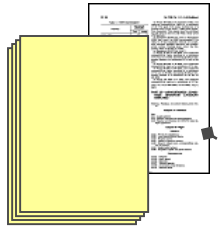


AC 25.1309 "Acceptable Means of Compliance to 25.13

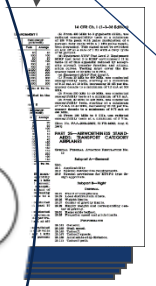


AC 20-115B

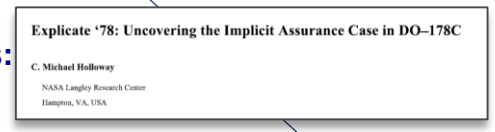


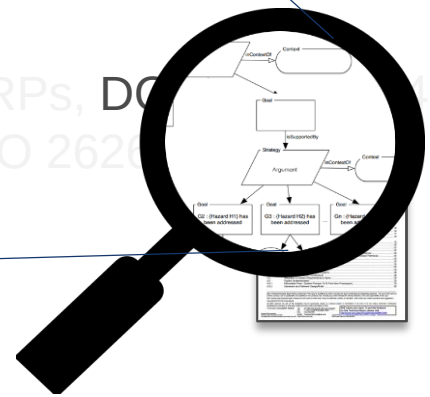


Aeronautics and Space



…niness standards:
…gory airplanes

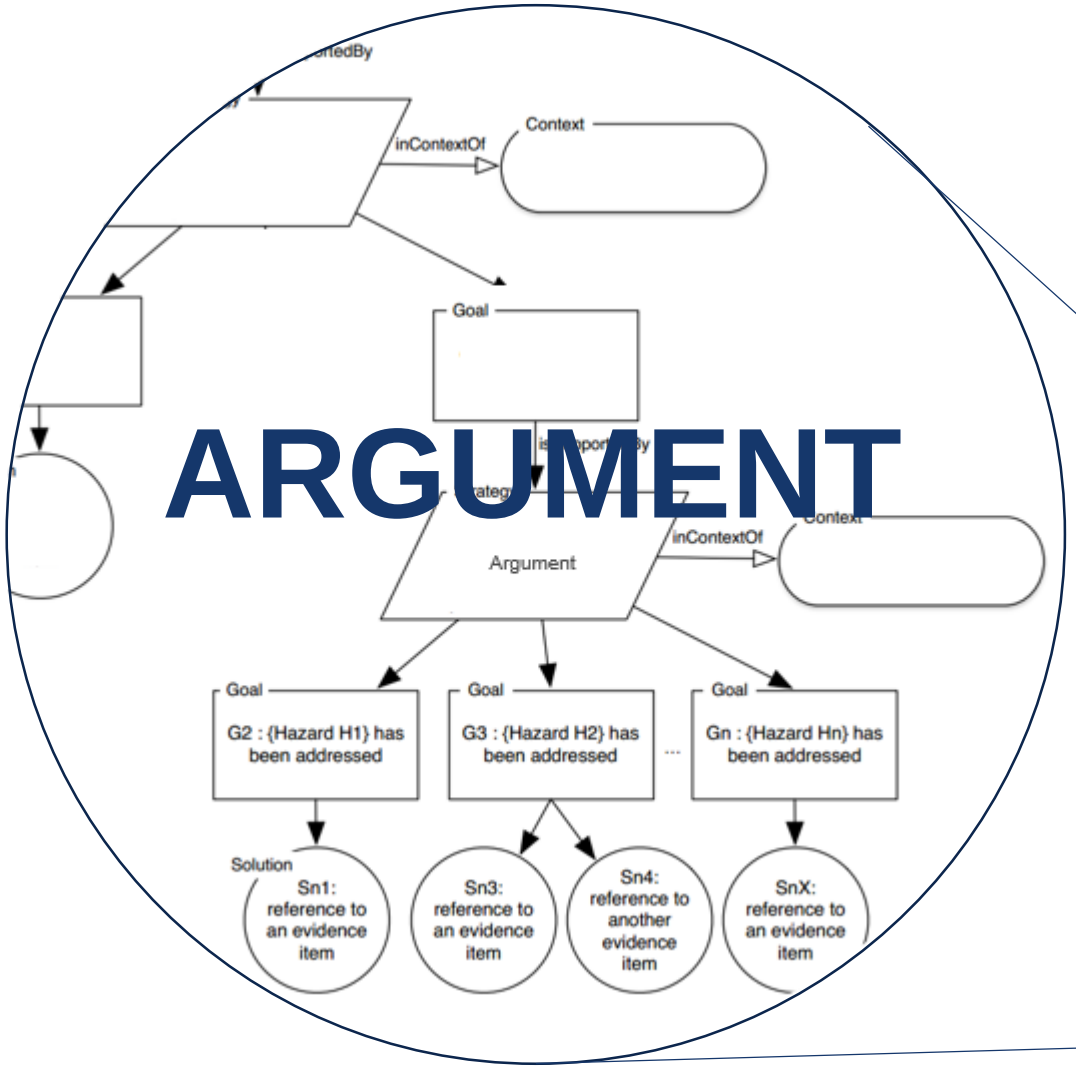


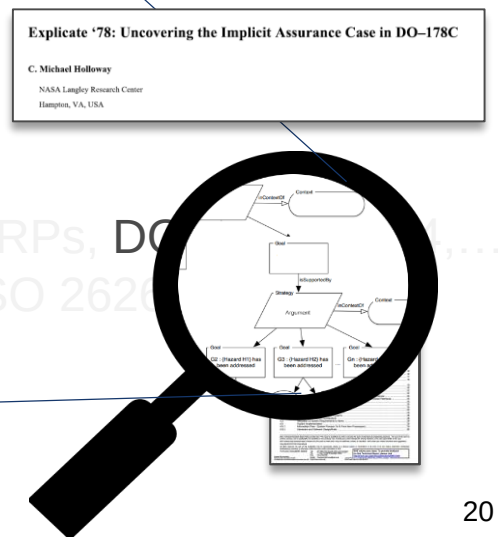


ARPs, DO… ISO 262…

Certification, aeronautics

- Requirements
- Determinism
- Observability
- Tracability
- Decomposition
- Experience
- etc





New computation paradigm, new problems…

Requirements

Determinism

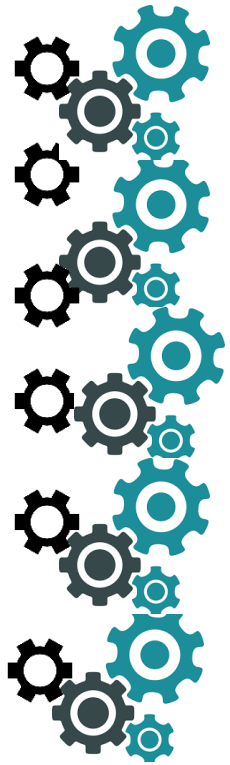
Observability

Tracability

Decomposition

Experience

etc



Requirements are sometimes **difficult to specify…** including environmental conditions (what are the “foreseeable operating conditions”?)

The behavior results from a learning process that is **statistical by nature** (find correlations)

The **role of data is crucial** during the learning phase, once the learning phase is completed. Effects of errors on those data is much more complicated to trace to the outputs…

Problems are solved “**globally**” (no decomposition, traceability)…

Models are usually **black boxes**…

“Usual” V&V methods needs to be “adapted” to ML (if possible), new methods need to be invented…

Approaches…

- **“Similarity analysis”**
- **“Backward analysis”**
- Inductive approach: **from faults to failures**
- “[Quasi] Deductive approach”: **high-level properties**

Approaches…

❑ **“Similarity analysis”**

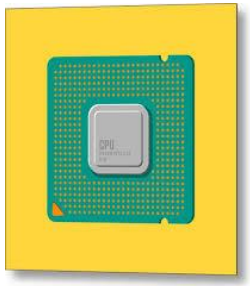
Aren’t we already facing similar situations?

Strong dependancy to data

Databases

Managed in an ad-hoc *manner*
Data are "engineered"
Data don’t express the behaviour

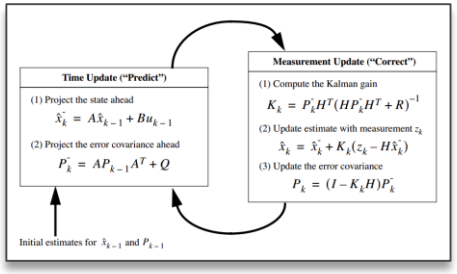
Epistemic uncertainty



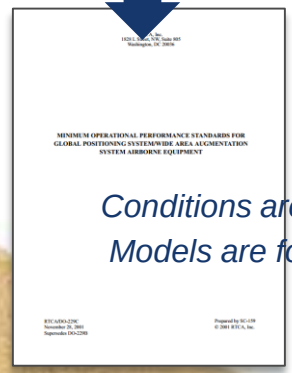
Complex processors (most systems?)

Managed in an ad-hoc *manner*
Covered by V&V (integration tests)
"Models" are available
Service experience

Stochastic uncertainty



Kalman filtering



Conditions are well-defined
Models are formal

Approaches…

Applied to ACAS Xu

- **"Backward analysis"**

Considering a ML technique providing a high-level confidence level, to which class of problems can we apply it ?

Reluplex: An Efficient SMT Solver for Verifying Deep Neural Networks

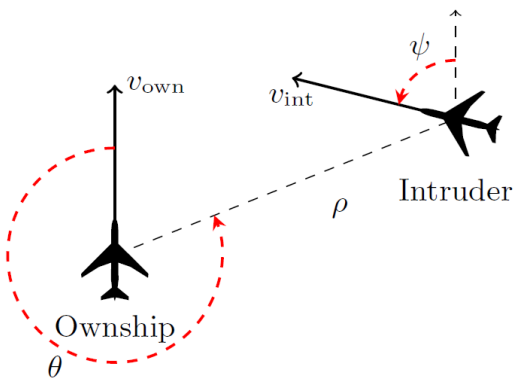
Guy Katz, Clark Barrett, David Dill, Kyle Julian and Mykel Kochenderfer

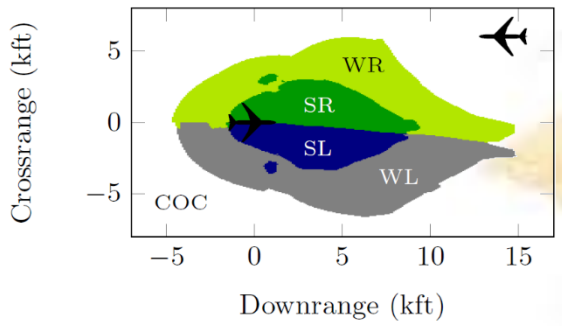
Stanford University, USA

19 mai 2017

Property ϕ_7.

- Description: If vertical separation is large, the network will never advise a strong turn.
- Tested on: $N_{1,9}$.
- Input constraints: $0 \leq \rho \leq 60760$, $-3.141592 \leq \theta \leq 3.141592$, $-3.141592 \leq \psi \leq 3.141592$, $100 \leq v_{\text{own}} \leq 1200$, $0 \leq v_{\text{int}} \leq 1200$.
- Desired output property: the scores for "strong right" and "strong left" are never the minimal scores.

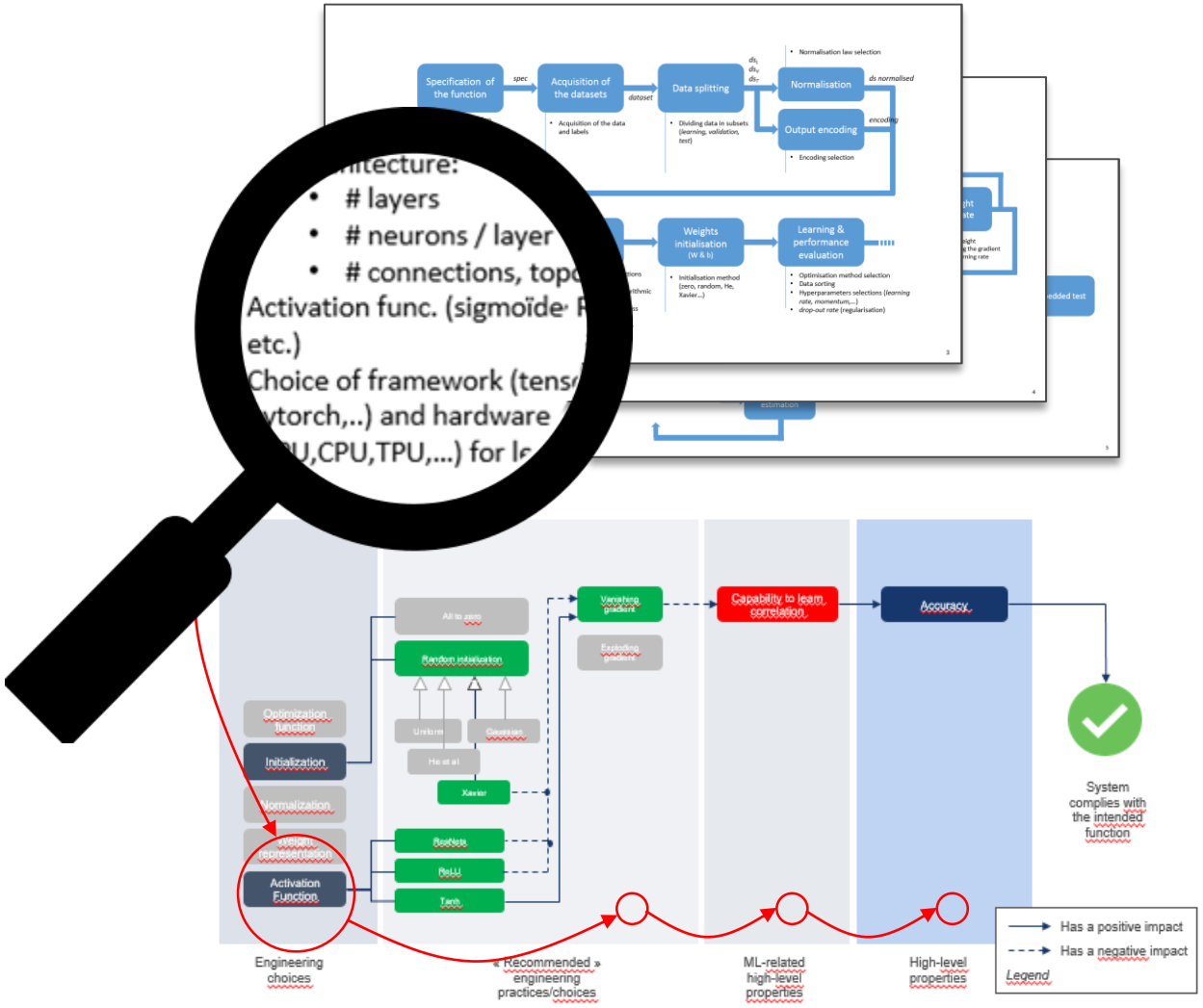




Approaches…

- Inductive approach: **from faults to failures**

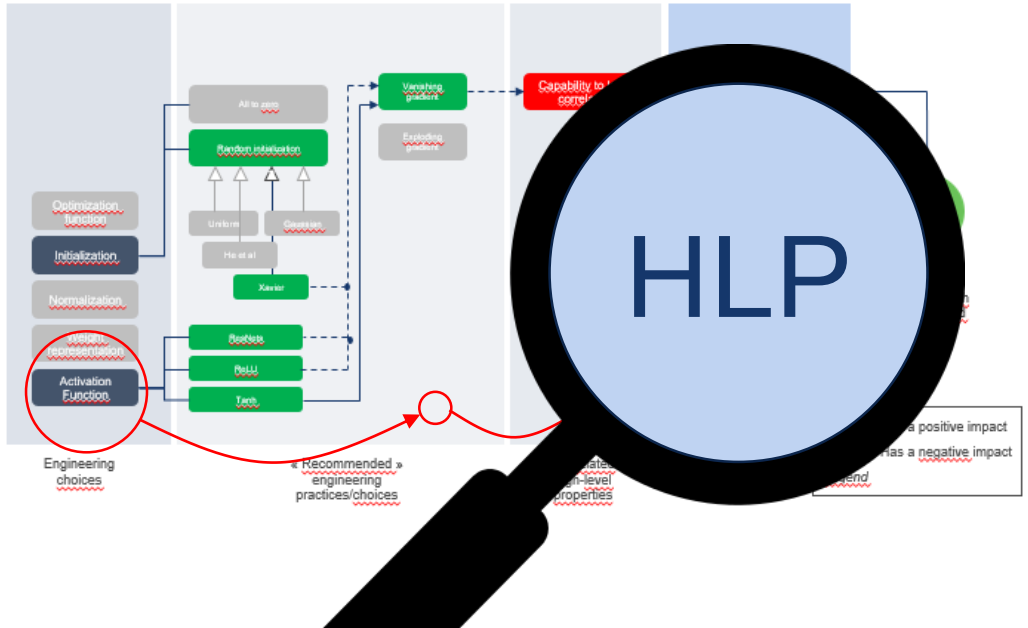
  Recipe:
  - Take a « typical » process,
  - Find where faults can be introduced (fault models)
  - Determine the effect of those faults on the top-level function
  - Determine the means to **prevent, detect** and **mitigate** the effects of those faults



Approaches…

- “[Quasi] Deductive approach”: **high-level properties**

"Properties that, if possessed, could have a positive impact on the capability to certify the ML-based system"

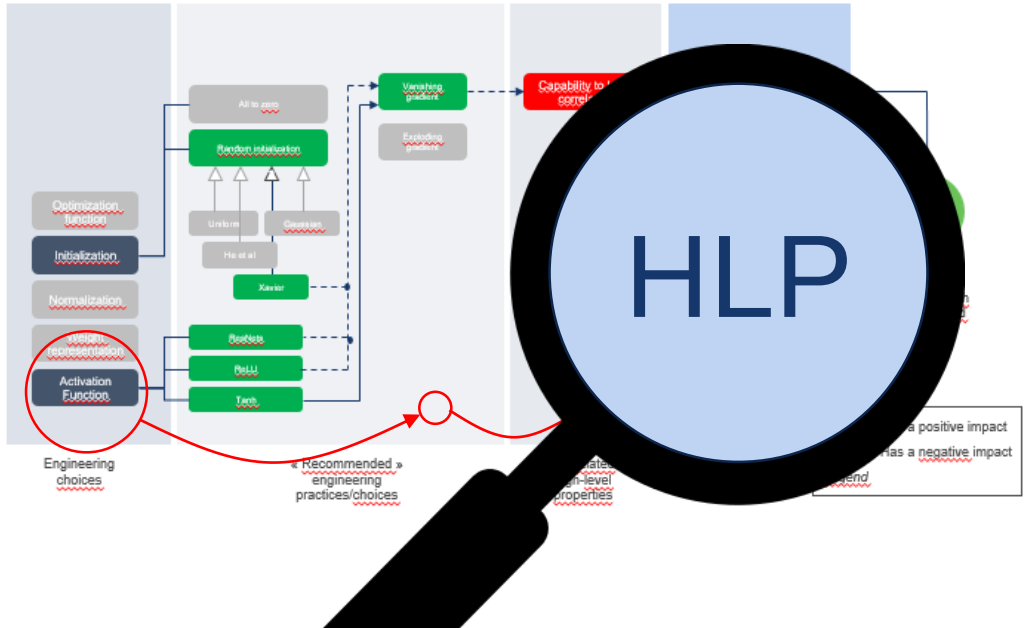


Approaches…

- “[Quasi] Deductive approach”: **high-level properties**

Specification issues

Probabilistic assessment

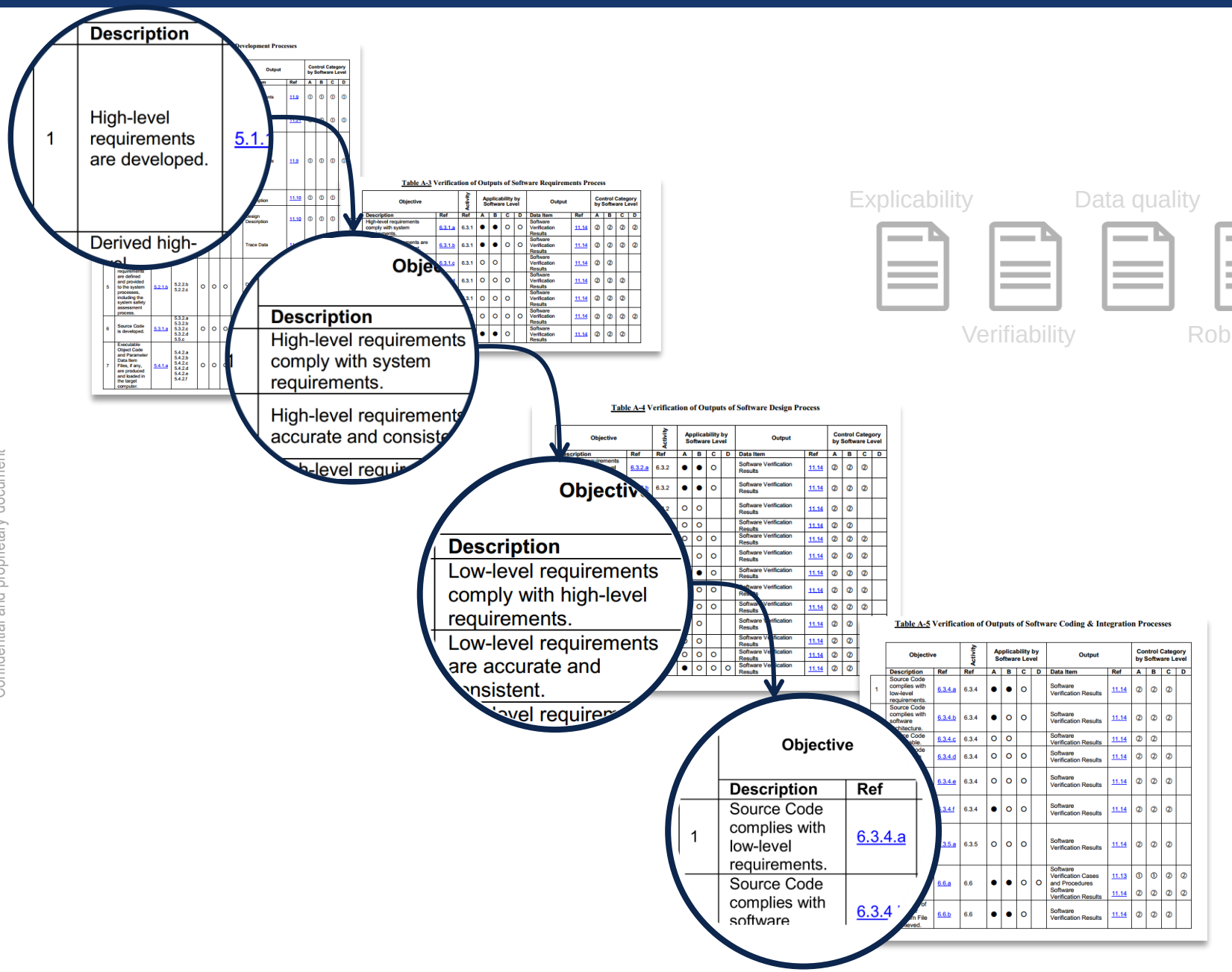


Challenges: focus on a few properties

Specification issues

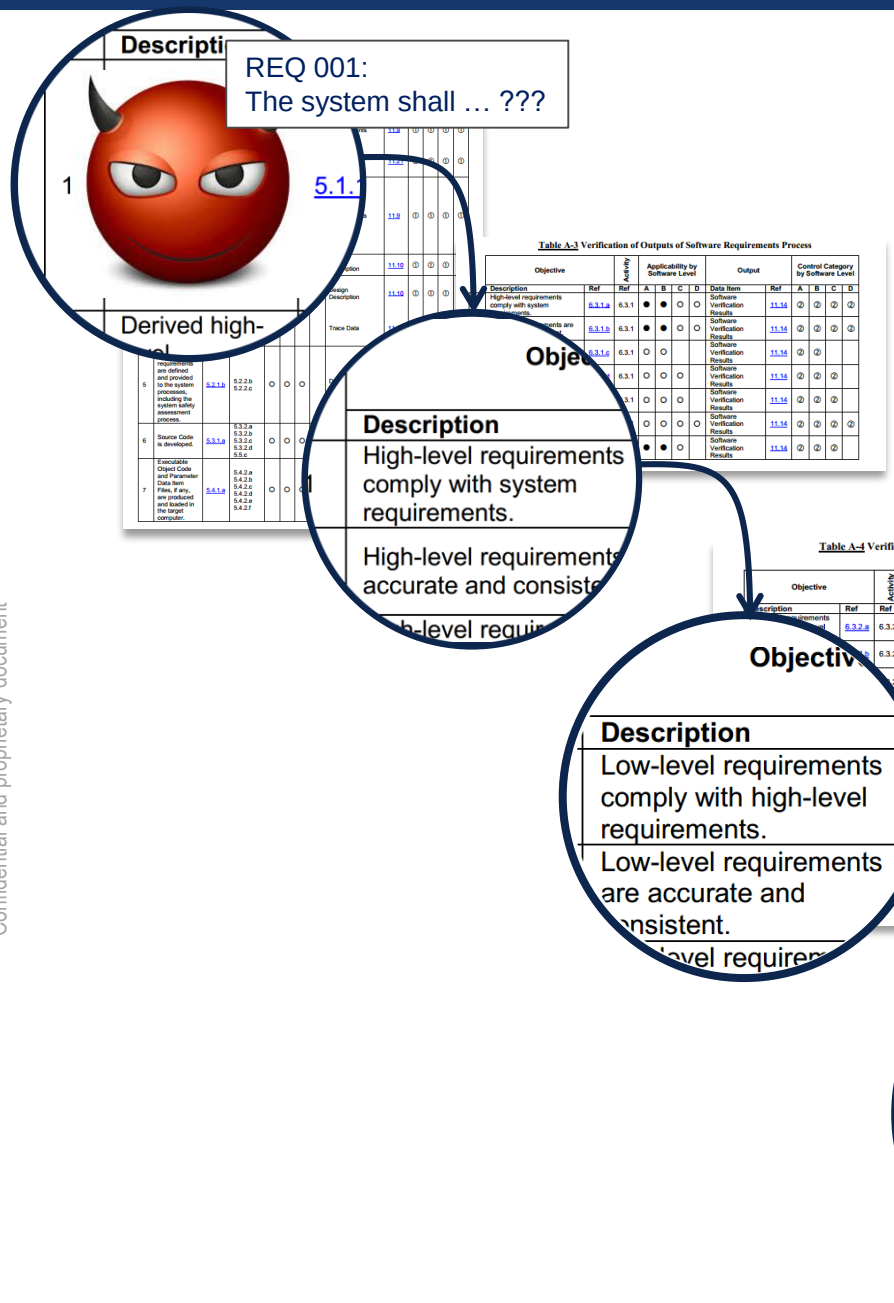
Probabilistic assessment

Challenges: focus on a few properties



Fundamental role of the specification **artifacts** in the design process

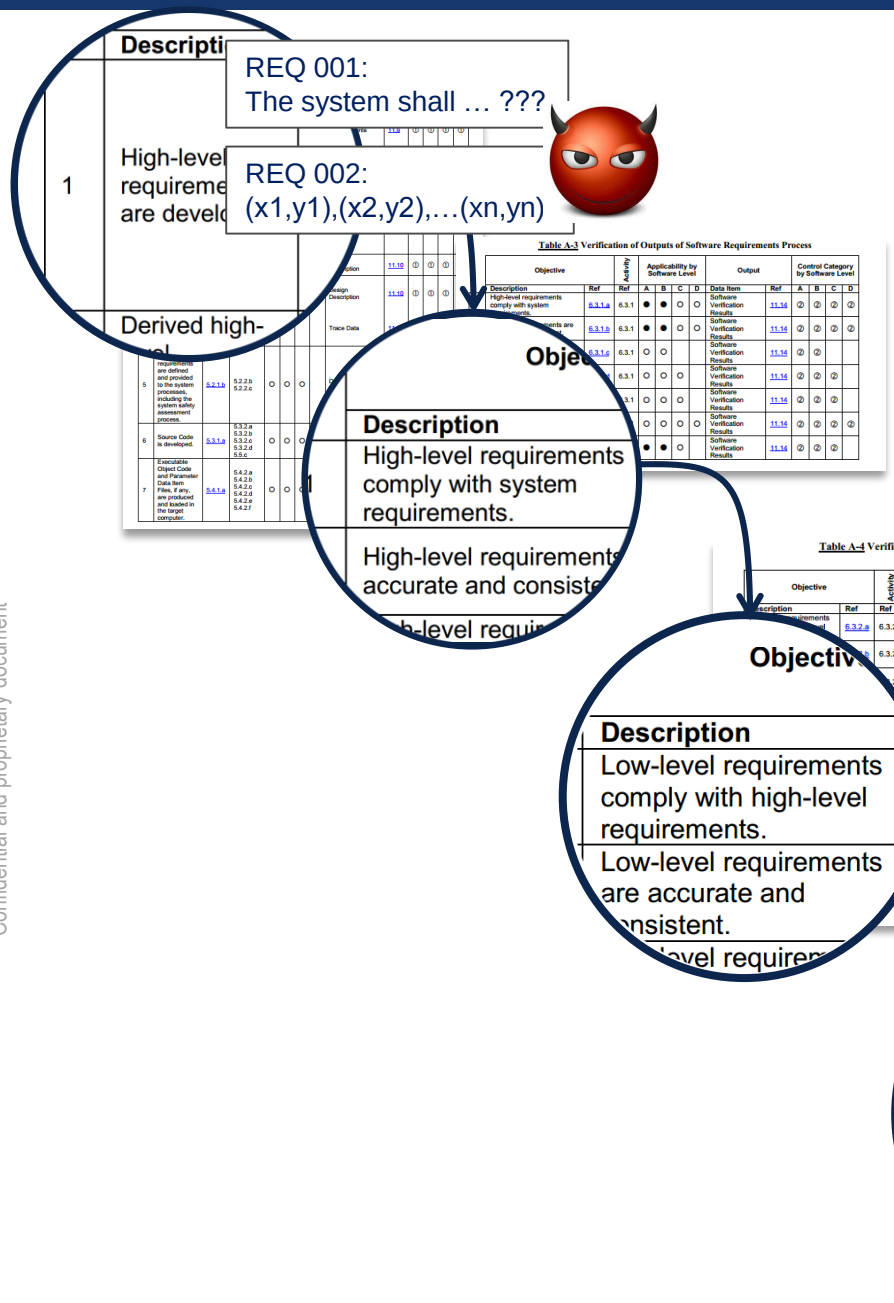
Challenges: focus on a few properties



Explicability Data quality Resilience … Verifiability Robustness

ML used in situations where other techniques [and human] have difficulties (e.g., *pedestrian detection*): **environmental conditions difficult to specify or control**

Challenges: focus on a few properties

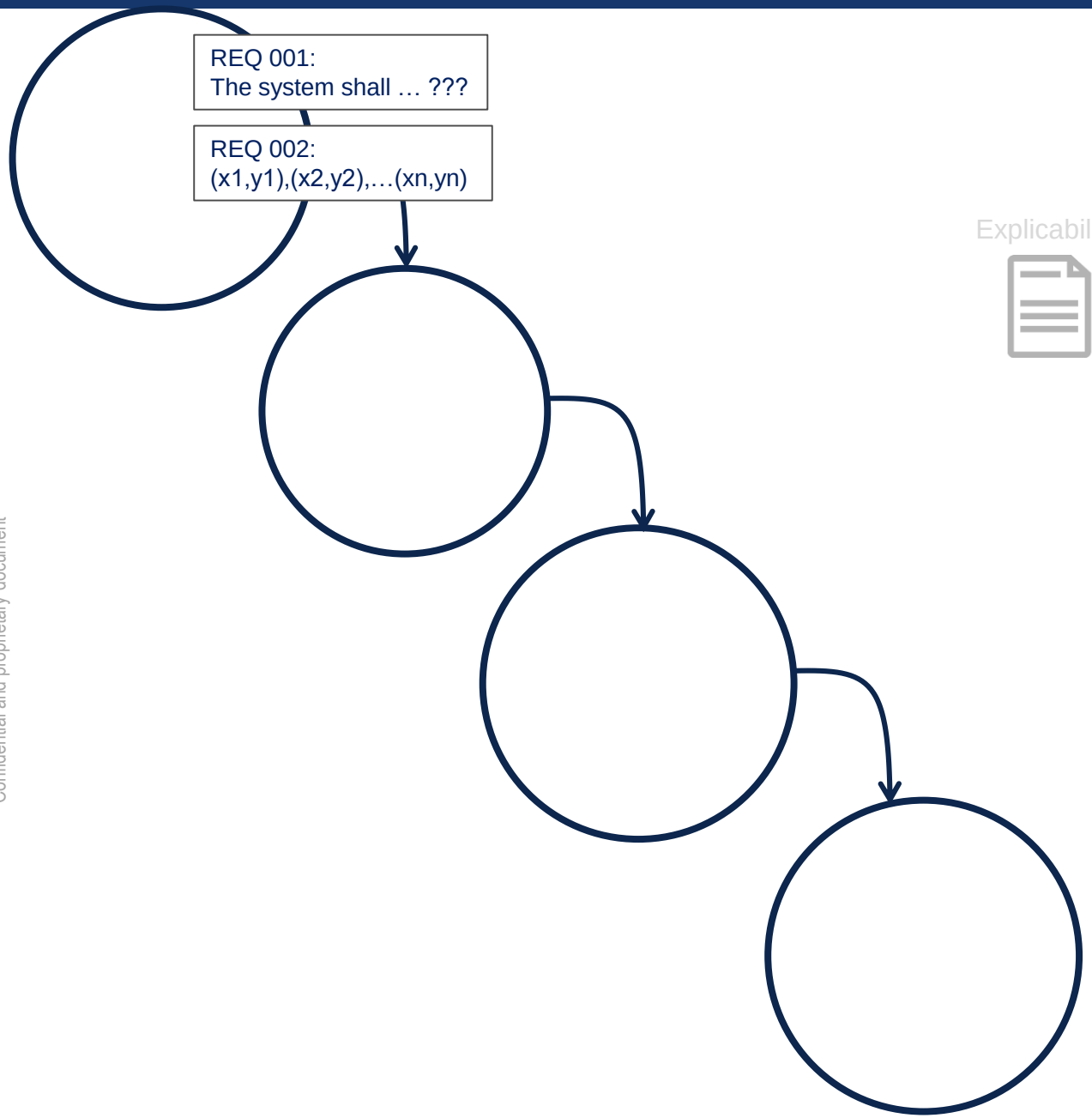


Learning dataset may be **part of the specification**

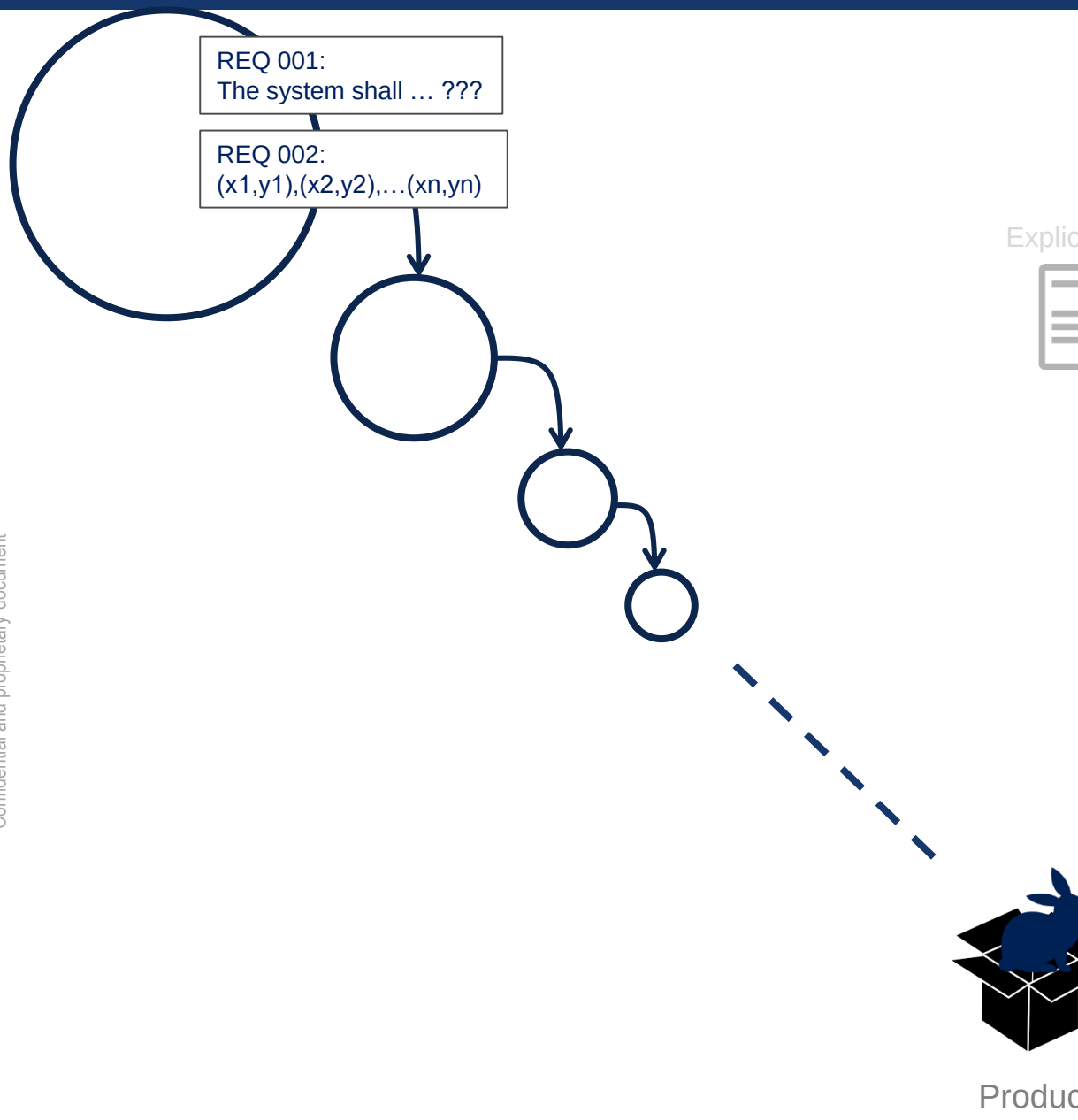
Challenges: focus on a few properties

REQ 001:
The system shall … ???

REQ 002:
(x1,y1),(x2,y2),…(xn,yn)



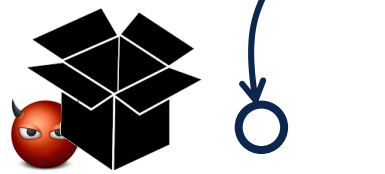
Challenges: focus on a few properties



Challenges: focus on a few properties

REQ 001:
The system shall … ???

REQ 002:
(x1,y1),(x2,y2),…(xn,yn)



ML inference
"machinery" (algorithms)

Classical design process

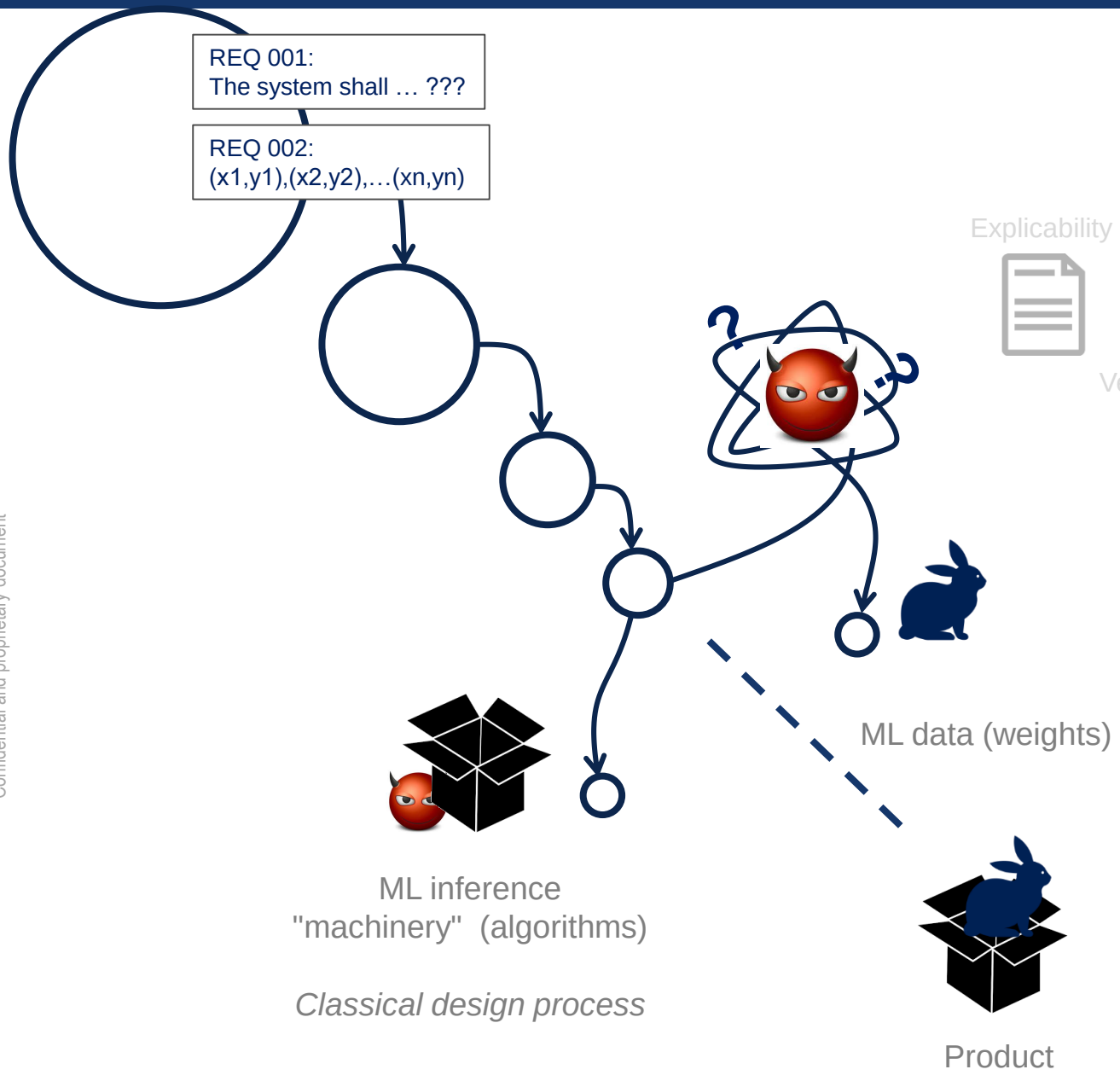
Product

Explicability Data quality Resilience

Verifiability Robustness

…

Challenges: focus on a few properties



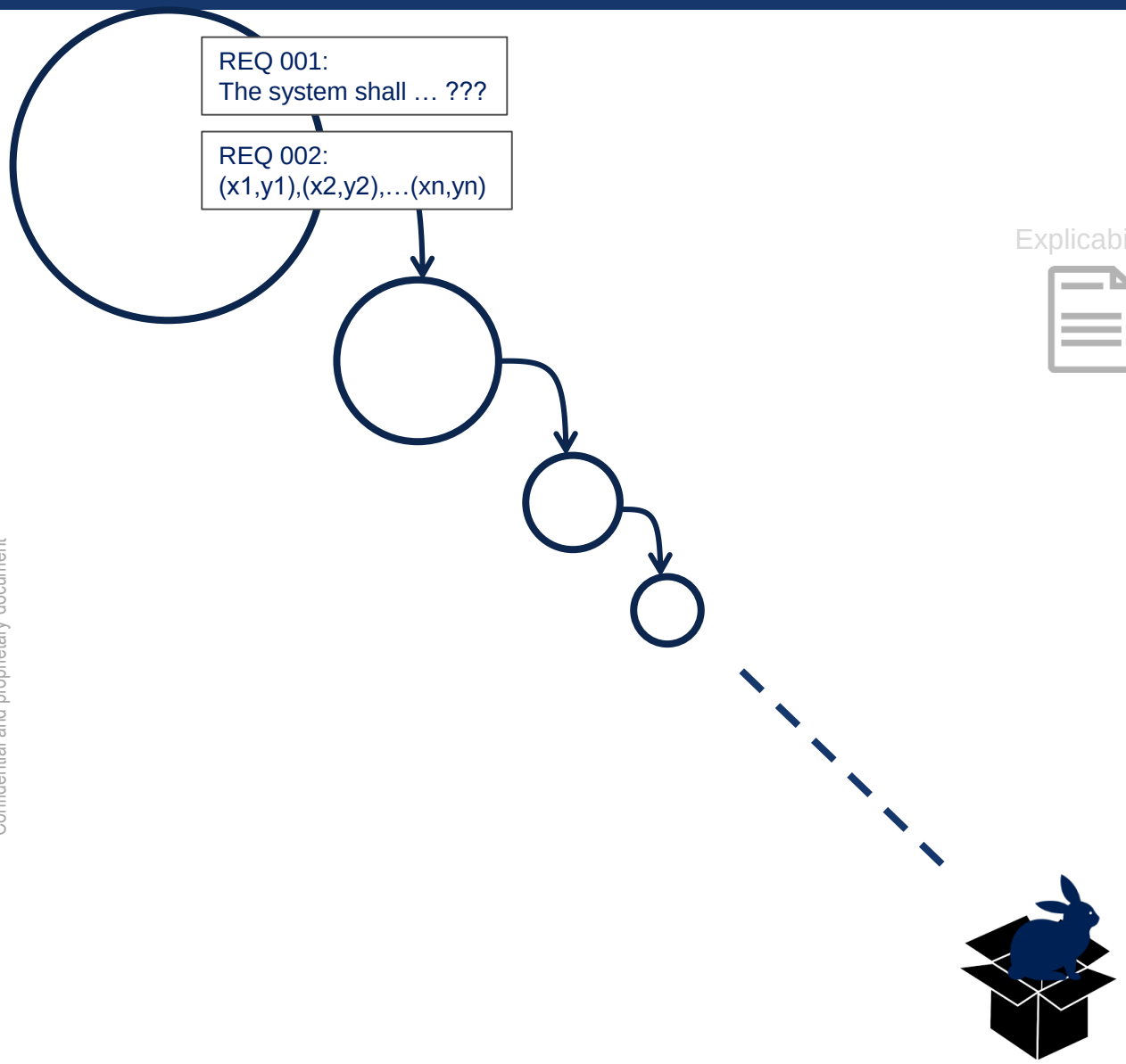
Lack of a **continuous refinement process**

Traceability to requirements is hard to ensure

Challenges: focus on a few properties

REQ 001:
The system shall … ???

REQ 002:
(x1,y1),(x2,y2),…(xn,yn)

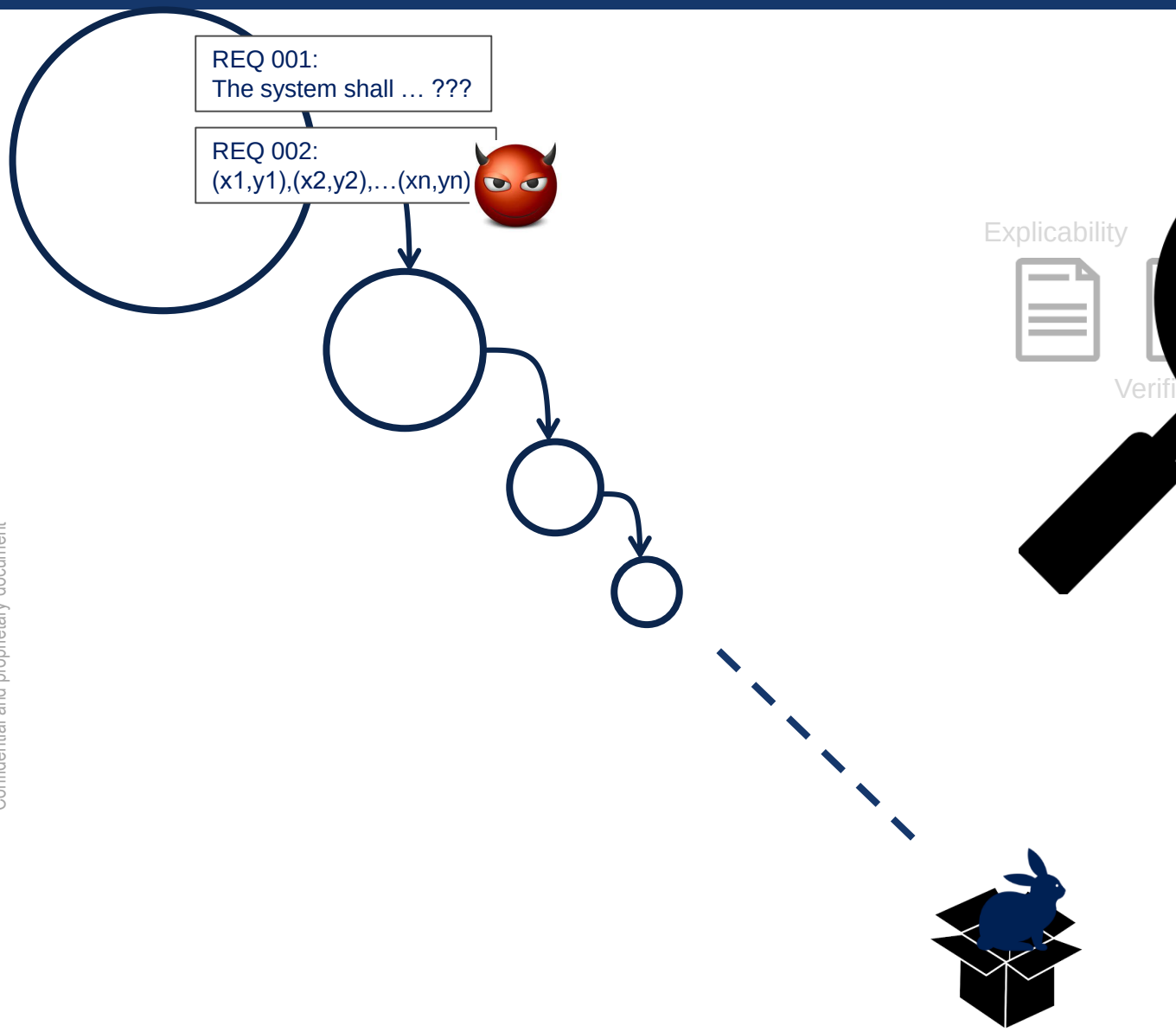


Product

Lack of a **continuous refinement process**

Traceability to requirements is hard to ensure

Challenges: focus on a few properties



Specification issues

Probabilistic assessment

Need for a "**Learning assurance process**"

Challenges: focus on a few properties

What explanations?
For **whom** and for **what purposes**?

Challenges: focus on a few properties

Explicability

Data quality

Resilience

…

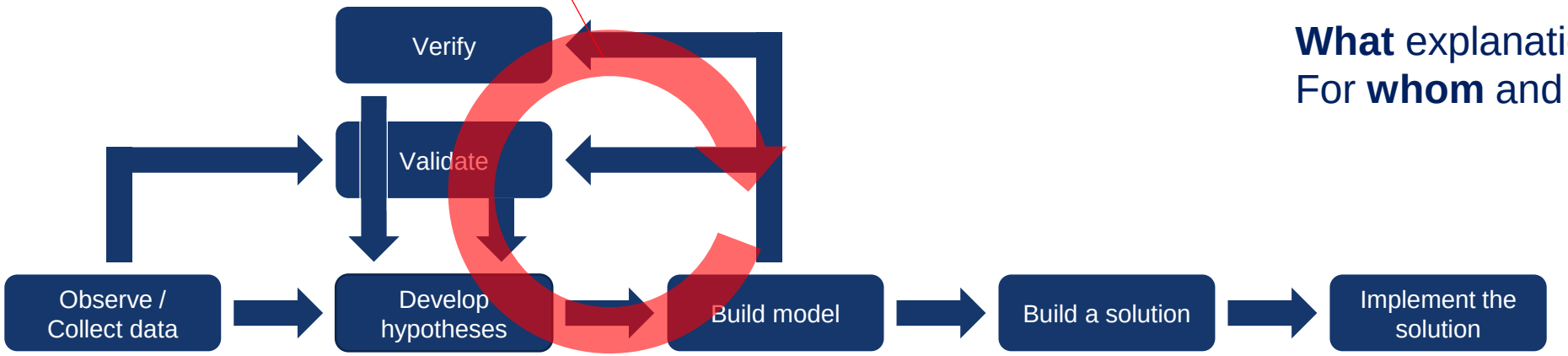
Specification issues

Probabilistic assessment

Robustness

Other…

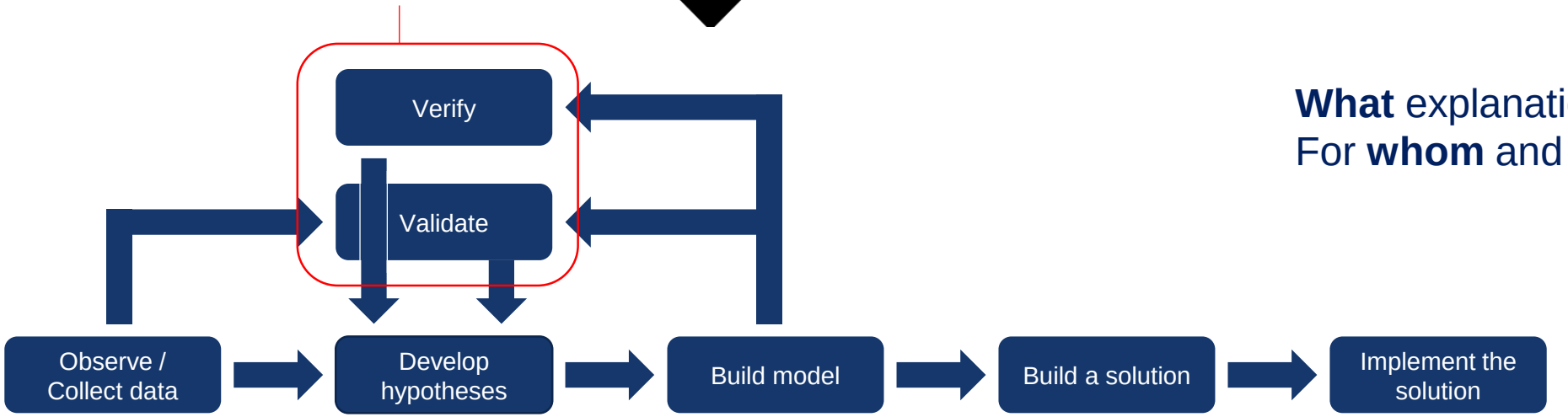
Hypothetico-deductive approach



What explanations?
For **whom** and for **what purposes**?

Challenges: focus on a few properties

The model is appropriate and is consistent.
It complies with the State of the Art.



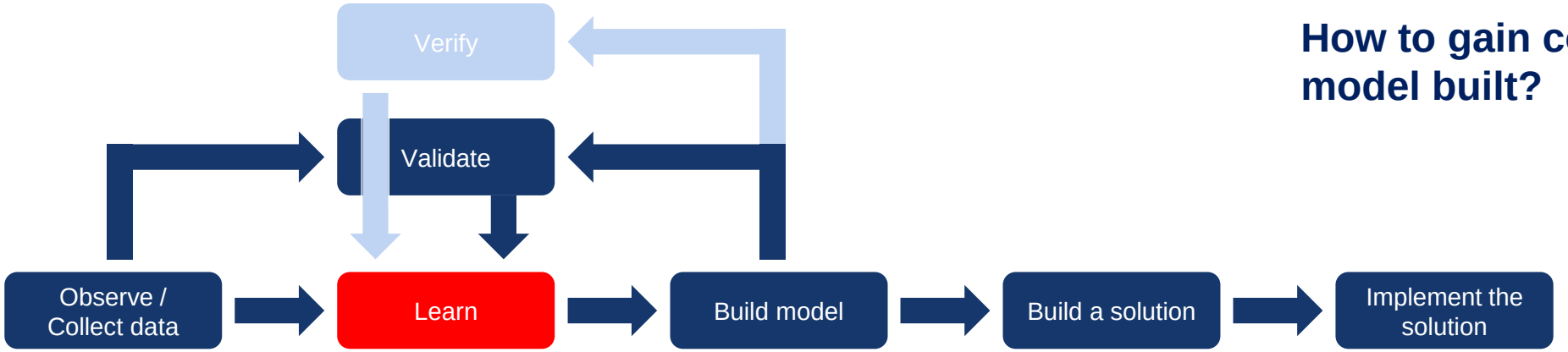
What explanations?
For **whom** and for **what purposes**?

Challenges: focus on a few properties

????

Specification issues

Probabilistic assessment



How to gain confidence on the model built?

Challenges: focus on a few properties

Can we **prove** an ML algorithm?

- *Are existing formal methods applicable to real-size problems?*

Can we **test** an ML algorithm?

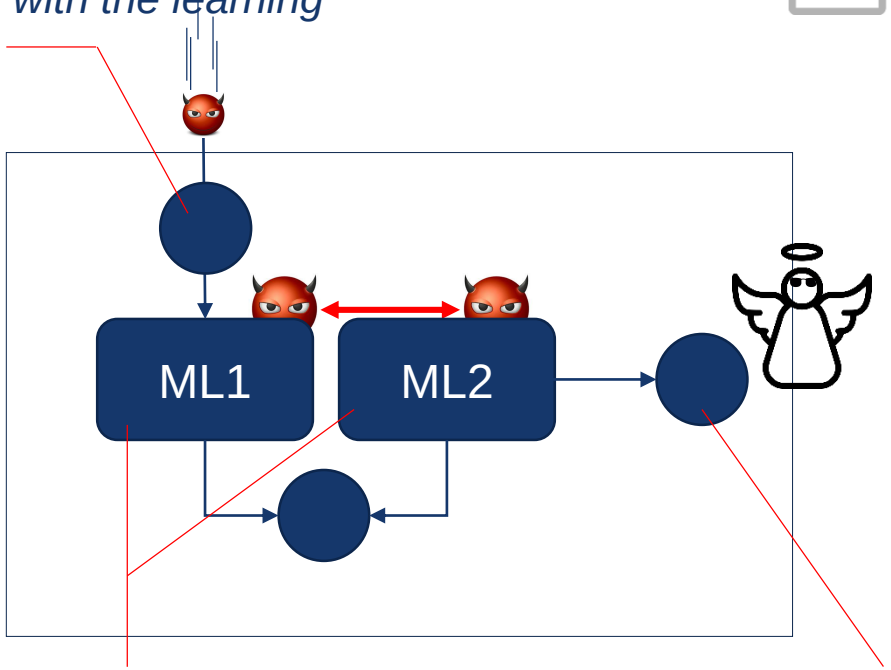
- *Where is the oracle? (see specification problem)*
- *Is massive testing massive enough? How to generate those tests...*
- *Equivalence classes?*
- *How to check innocuity?*

How to gain confidence on the model built?

How can we demonstrate compliance to the requirements?

Challenges: focus on a few properties

How to ensure that the actual conditions comply with the learning conditions?



Can we apply classical “system-level” architectural recipes?

How to diversify (and avoid common causes)?

How to monitor a ML system?

Challenges: focus on a few properties

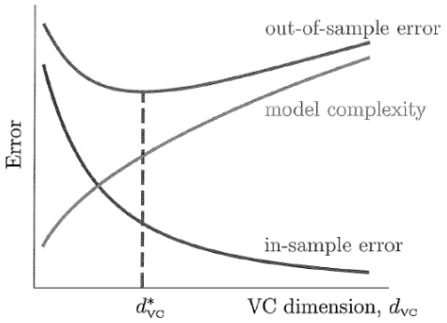
- *Learning is a statistical process*
- *Some results about probabilistic assessment*

Vapnik and Chervonenkis dimension

$$P\left(Err_{op} \leq Err_{train} + \delta\right) \leq 1 - \varepsilon \; where \; \delta = f(VC, N, \varepsilon)$$

Size of the datasetf

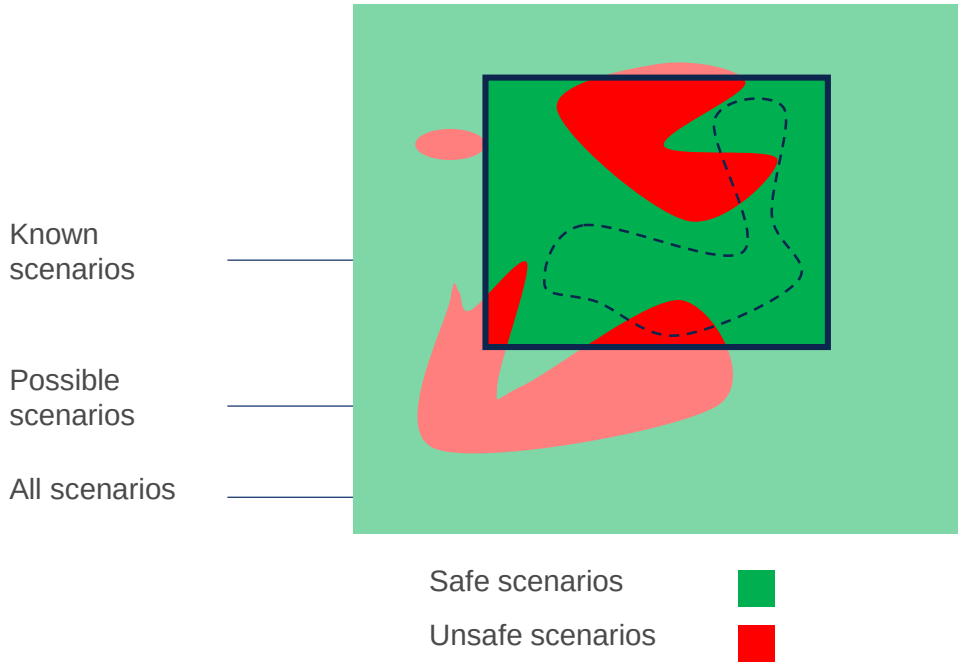
$10^{-5} \longrightarrow \delta \leq \sqrt{\frac{VC}{N}}$ ← 10^5 for DNN → 10^{15}



Can we rely on a probabilistic assessment of dependability?

Challenges: focus on a few properties

- *Learning is a statistical process*
- *Some results about probabilistic assessment*
- *The nature of uncertainty*



Can we rely on a probabilistic assessment of dependability?

Challenges: focus on a few properties

- *Learning is a statistical process*
- *Some results about probabilistic assessment*
- *The nature of uncertainty*

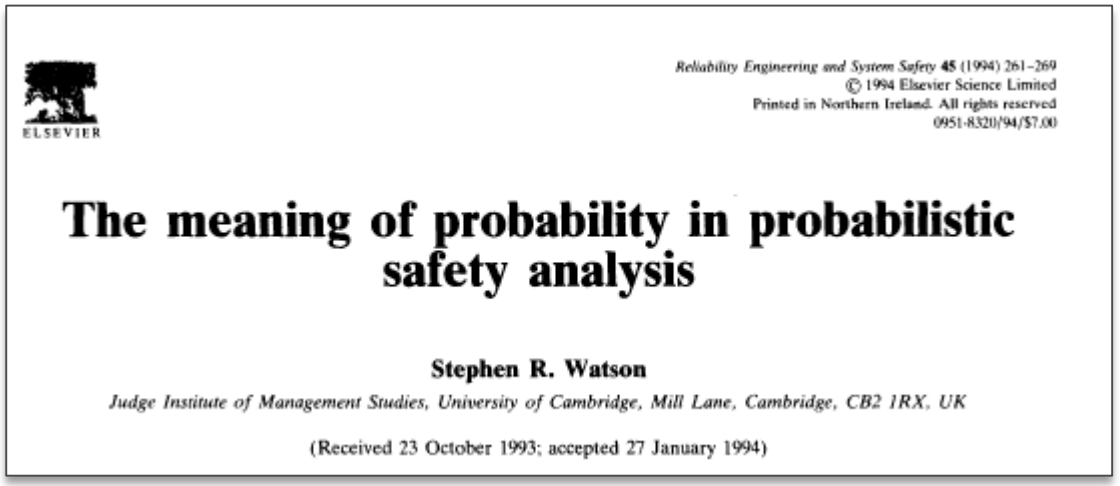
Aleatoric uncertainty?

Epistemic uncertainty?

Can we rely on a probabilistic assessment of dependability?

Challenges: focus on a few properties

- *Learning is a statistical process*
- *Some results about probabilistic assessment*
- *The nature of uncertainty*
- *The various interpretations of probabilities…*
  - *Classical, frequentist, degree of belief?*



ELSEVIER

Reliability Engineering and System Safety **45** (1994) 261–269
© 1994 Elsevier Science Limited
Printed in Northern Ireland. All rights reserved
0951-8320/94/$7.00

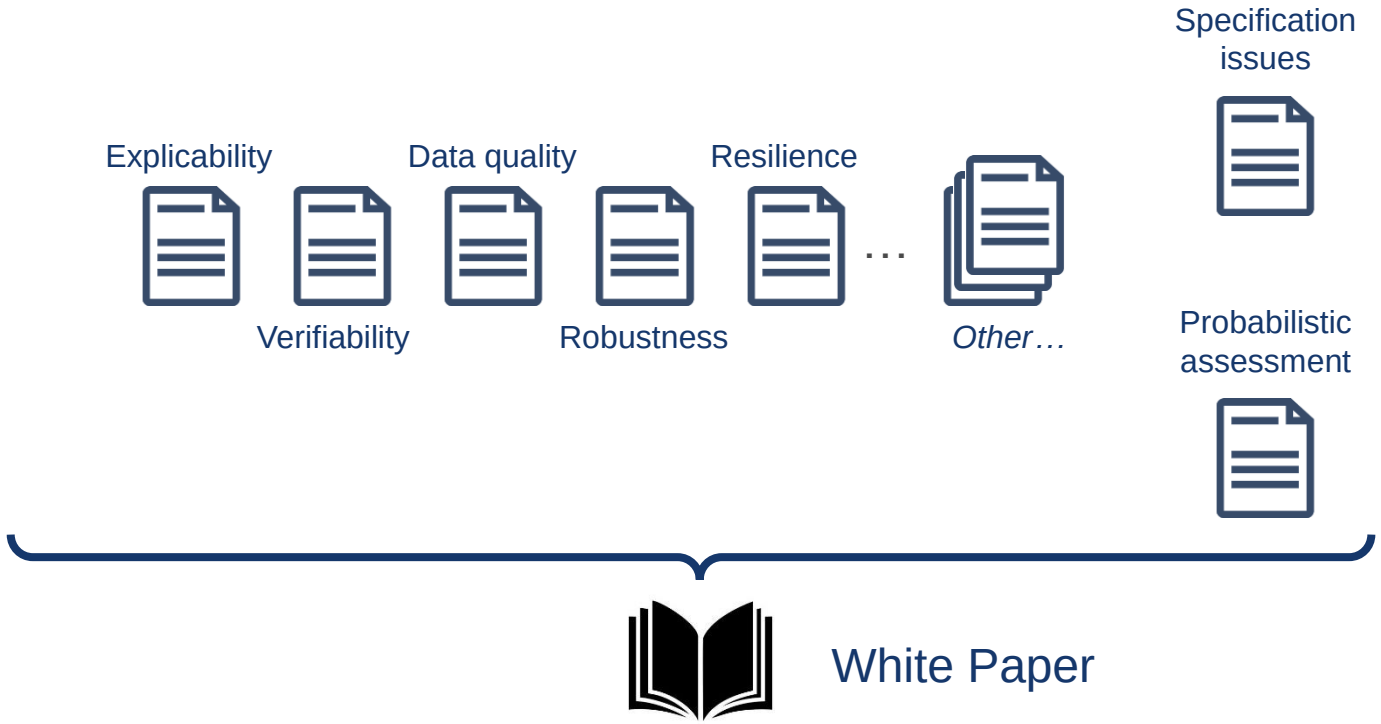
The meaning of probability in probabilistic safety analysis

Stephen R. Watson
Judge Institute of Management Studies, University of Cambridge, Mill Lane, Cambridge, CB2 1RX, UK

(Received 23 October 1993; accepted 27 January 1994)

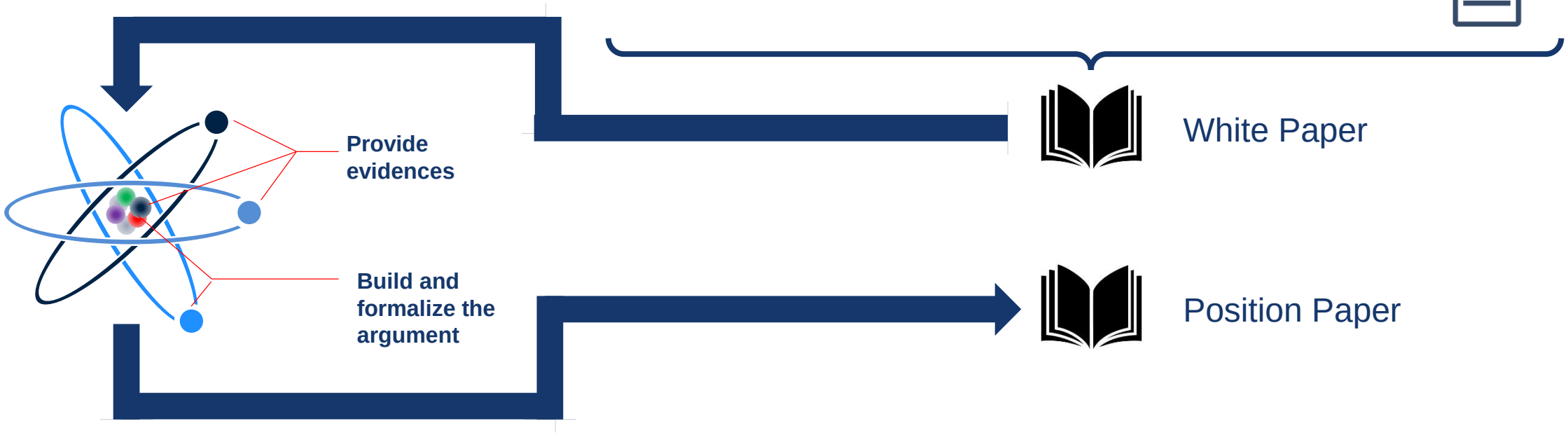
Can we rely on a probabilistic assessment of dependability?

We are we now…



And ware we going next?

- On two use cases
  - ACAS Xu
  - *Image processing app. TBD*



DEEL
DEpendable & Explainable Learning

THANKS FOR YOU ATTENTION

IVADO IRT SAINT EXUPERY CRIAQ ANITI ARTIFICIAL & NATURAL INTELLIGENCE TOULOUSE INSTITUTE