

ANITI

ARTIFICIAL & NATURAL INTELLIGENCE
TOULOUSE INSTITUTE



THEME 6 – Certifiable AI: safe design and embeddability



Université
Fédérale
Toulouse
Midi-Pyrénées

AGENDA (20min presentation + 10min questions)



Members: 1 min

Definition/scientific perimeter of the theme & threads: 8 min

On going work: 5min

Highlights & main results : 2 min

Scientific animation of the theme: 4min

MEMBERS



Chairs: 3 chairs and part of DEEL

Serge Gratton:

- **people involved:** Pierre Boudier, Alfredo Buttari, Serge Gratton
- **ANITI Resources:** 1 PhD Cifre Atos Théo Beuzeville

Joao Marques Silva:

- **people involved:** Emmanuel Hebrard, Joao Marques Silva

Claire Pagetti:

- **people involved:** Mohammed Belcaid (MAD), Kevin Delmas, Jérémie Guiochet, Charles Lesire-Cabaniols, Claire Pagetti, Aurélien Plyer
- **ANITI Resources:** 1 PhD with NXP Iban Guinebert, 1 PhD with Airbus, 1 post doc Joris Guerin

DEEL project (mission certification AI):



- **People involved :**
Franck Mamalet, Eric Jenn, Grégory Flandin, Sébastien Gerchinovitz (IRT),
Claire Pagetti, Kevin Delmas (Onera)
- **DEEL ressources:**
Industrial MàD (20%) from Airbus, Apsys, Continental, DGA, EdF, Renault,
Safran, Scalian, SNCF, Thales
Core Team MàD and IRT (>50%)
- **Invited speakers during workshop:**
M. Serrurier (IRIT), JM Loubes, F. Malgouyres (IMT) , E. Pauwels (IRIT), J.
Guiochet (LAAS), Fabrice Gamboa (IMT), D. Bertoin (IRT)

« CERTIFICATION IA » WORKGROUP

11 partners in 4 domains:

- Aero, Automotive, Railway, Energy

Two-days working group meetings per month at
IRT St Exupery



Main objectives:

- « Acculturation » (IA<->certification/inter domain)
- Create knowledge (input for certification groups, for instance a White Paper on AI Certification)
- Provide elements of confidence for certification (based on use cases and the core team)



White Paper

Machine Learning
in Certified Systems



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

DIFFUSION RESTREINTE/RESTRICTED DIFFUSION

Context: certification activities

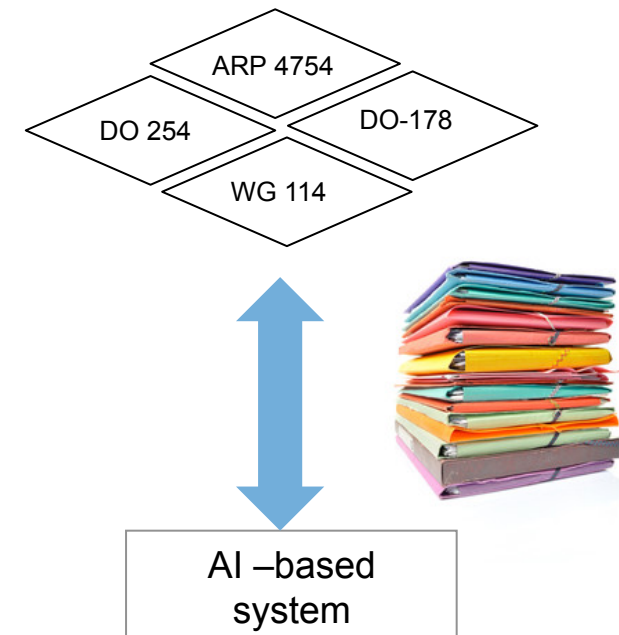
Certification:

- evaluation of an **argumentation**, to convince that a system (i.e., its architecture, its settings, including mitigation means. . .) satisfies **certification objectives** (expressed with AMC standards)

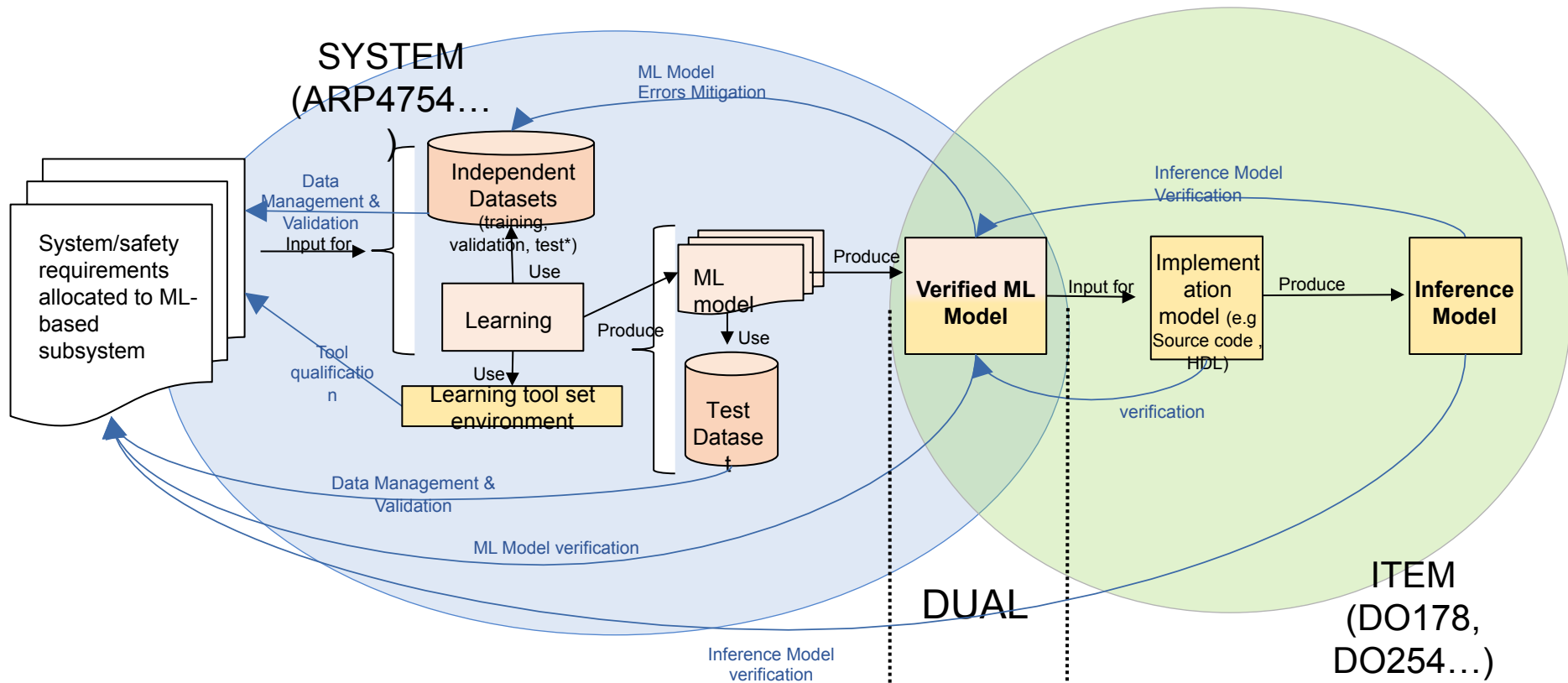
Difficulties :

- Existing standards are inapplicable [BCM+15]
 - Data oriented specification

[BCM+15] Siddhartha Bhattacharyya, Darren Cofer, David J. Musliner, Joseph Mueller, and Eric Engstrom. Certification considerations for adaptive systems. Technical Report NASA, 2015



EUROCAE WG114 current vision

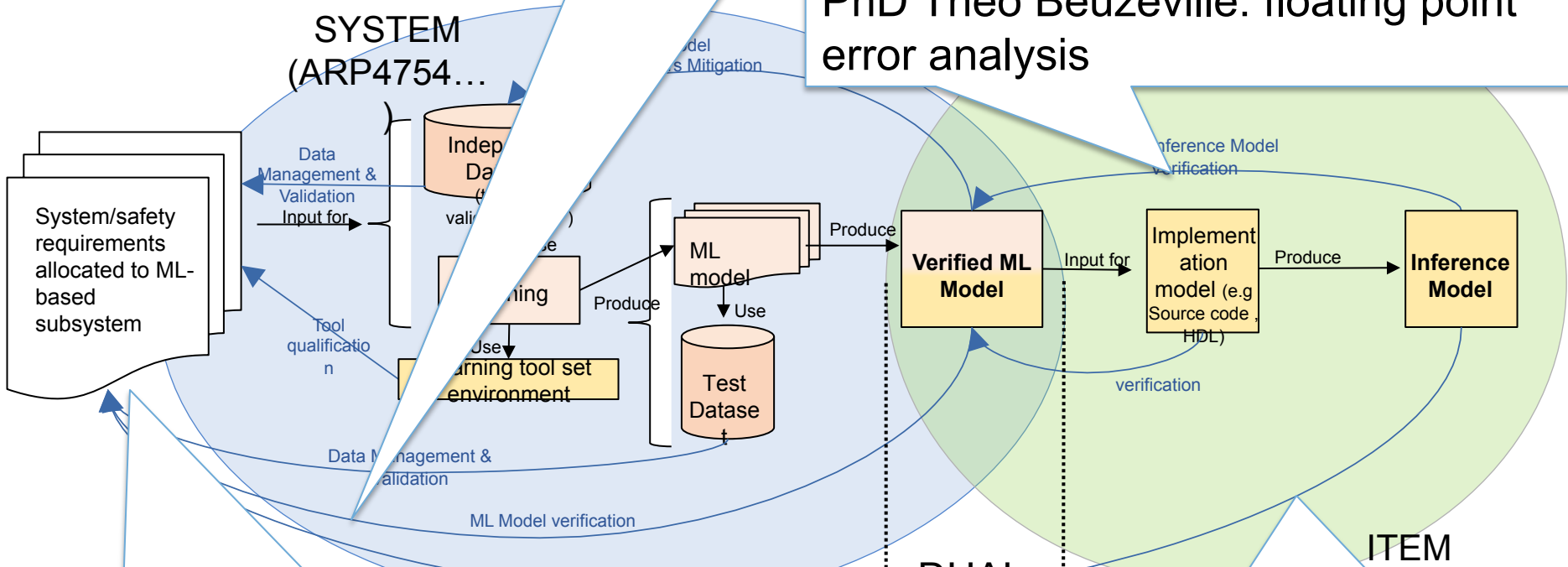


EUROCAE WG114 current vision



Mission certif DEEL / Marques Silva

PhD Théo Beuzeville: floating point error analysis



Mission certif DEEL / Post Doc Joris Guerin: runtime verification

DUAL PhD Iban Guinebert: hardware failure impact

IP Certifiable AI



Title 6.1 Embeddable AI architecture

Objectives:

The problem is to provide physical /software constraints around ML or deep learning algorithms that will yield results consistent with certification requirements.

The contributions will cover several types of requirements: managing off-line floating point errors, ensuring the capacity to assess a WCET and detecting at run-time hardware failures.

Chairs

Gratton
Pagetti

Partners

Airbus
NXP

Resources

2
3

Challenges

- Mastering COTS/SOC hardware

Resources

- 1MS - 1 PhD of A. Buttari - Théo Beuzeville (Error analysis for Deep Machine Learning methods and tools) - (Cifre Atos)
- 1 MS - 1 PhD of C. Pagetti with NXP - Iban Guinebert (Methodology for integrating neural network IP in a chip with safety assurance)- (Région-NXP)
- 1 MS - 1 PhD of C. Pagetti with Airbus (Certified programming framework for machine learning applications)

Tools/Techniques

- Execution model definition
- Deterministic and probabilistic error analysis

Applications/Use cases

- important for critical systems

Error analysis for ML methods and tools



$$fl(a \otimes b) = (a \otimes b)(1 + \delta), \delta \leq u$$

Precision	bits	u
BFLOAT16	16	$\sim 10^{-3}$
FP16 (half)	16	$\sim 10^{-4}$
FP32 (single)	32	$\sim 10^{-8}$
FP64 (double)	64	$\sim 10^{-16}$

In critical systems the overall error may depend on differences in

- hardware
- choice of precision in training vs inference
- implementation of ML operators (e.g. FFT vs Winograd for convolution)

Objective: analyze and quantify error accumulation in DML methods

- Deterministic approach to bound the worst-case error
- **Probabilistic** approach: assume that δ are independent random variables of mean zero \rightarrow sharper (**several orders of magnitude**) practical error bounds

Challenge: define a general analysis framework that takes into account DML features such as composition of layers, nonlinearity of activation functions etc.

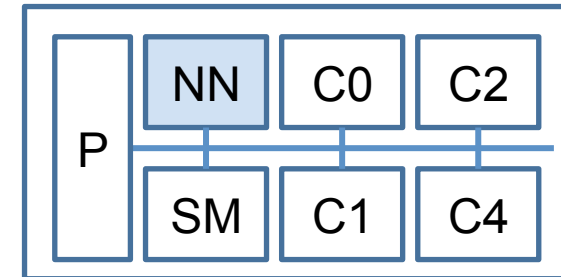
Impact:

- Fine tune precision in order to minimize CPU, memory and energy **consumption** without affecting **accuracy** and **portability**
- Guarantee **embeddability** and portability of DML methods/tools

Methodology for integrating neural network IP in a chip with safety assurance

New processor architecture:

- NN (neural network) HW IP in multi-processor chip for automotive
- SM (Safety Manager) in charge of supervising the platform
- Compliance with ISO26262



Objectives:

- Identification of the NN hardware random failures, the associated failure modes and the effect at the platform level
- How the safety manager should be modified and extended to integrate the NN component
- Definition of a generic safety methodology to assess the NN within the platform.

IP Certifiable AI

Title 6.2 Model level selection & verification

Objectives:

The objective is to offer a high level of safety for machine learning based systems. The approach consists in offering several verification means of the trained model and show that it is compliant with the system requirements.

The contributions will cover any argument that provides assurance on the model. The first approach concerns the formal verification of the trained model and will incorporate: the development of new verification tools, the definition of verification strategies and the combination of simulation and formal verification when the verification only is insufficient.

The second approach consists in defining alternative techniques when formal verification cannot be applied. An example of such alternative is the definition good learning practices that will provide some confidence on the model.

Challenges

- Lack of formal specification, lack of formal properties to be verified, large size systems





Tools/Techniques

- various verification tools: Mixed integer linear programming, SAT solvers (binarized Neural Nets)
- Good practices, assurance cases

Applications/Use cases

- important for critical systems



Chairs	Partners	Resources
Marques Silva Pagetti	CS	15 
DEEL mission certif	DEEL Mission certification partners	  

Resources
<ul style="list-style-type: none">• 1MAD: CS (Pagetti: Algorithms for certifiable Deep Learning)• Mission certif sprint ACAS and sprint vision• DEEL MS: Eduardo Dadalto Gomes



IP Certifiable AI

Title 6.3 System level assurance

Objectives:

The objective is to offer guaranties and assurance at the system level, the system including some ML based component.

At the system level, it is mandatory to analyse whether the overall system fulfills the safety objectives. In particular, the safety assessment consists in making dysfunctional model and evaluate the qualitative (cut sets) and quantitative (e.g. reliability) attributes of the system. This is a specific work of the DEEL mission certif.

Runtime verification: An independent safe system is embedded with the ML functions to observe them and detect whether a dangerous situation occurs. Such a solution is classical in safety critical systems. The novelty comes from the specification of the monitor as there does not exist a formal specification of the main system to be observed.

[Link with thread 11.7](#)

Challenges





Tools/Techniques

- Safety assessment
- Runtime verification

Applications/Use cases

- important for critical systems



Chairs	Partners	Resources
Pagetti	DEEL Mission certification partners	5 
DEEL mission certif		1 
		
		

Resources
<ul style="list-style-type: none">• 1 post doc Pagetti: runtime verification for critical machine learning applications – Joris Guerin• Mission certif sprint probabilistic assessment



Runtime Verification for Critical Machine Learning Applications

New “threats” of ML component in critical system:

- Distributional shift, concept drift, domain shift
- Corner case, Adversarial

Problem: Issues to formally demonstrate ML threats acceptability

Idea: Enable **fault tolerance** with runtime monitoring detecting **safety impact of slight modifications** (e.g. adversaries generation)

Objectives:

- Specification & design of runtime monitor
- Verification of detection properties on ML + Monitor
- Implementation and assessment on UAV Emergency Landing System

On-going work



On-going collaboration between chairs:

- DEEL mission certif and Pagetti
- Alami and Pagetti
- DEEL / CANADA

On-going collaboration with ANITI Industrial partners

- Airbus, Apsys, ATOS, Continental, CS, DGA, EDF, NXP, Renault, Safran, Scalian, SNCF, Thales

On-going ANITI Phd & Post doc

- Théo Beuzeville (submission to ANRT soon)
- Iban Guinebert start Nov 1st 2020
- Joris Guerin start Nov 1st 2020

On-going work



Standardization group participation

- EUROCAE WG114, SOTIF
- Regular discussion with EASA

On-going collaboration with external projects (national, EU, industry...)

- Cifre with Collins

Highlight & main results



Publications

- Identifying Challenges to the Certification of Machine Learning for Safety Critical Systems. Eric Jenn et al. ERTS 2020. February 2020
- White paper Mission Certif DEEL: Machine Learning in Certified Systems. IRT Saint Exupéry et al. June 2020.
- Ensuring Dataset Quality for Machine Learning Certification. Eric Jenn et al. WoSoCer. 2020
- Challenges in certification of computer vision based systems for civil aeronautics. Frédéric Boniol et al. Aerospace Lab. November 2020.



Scientific event organization & participation

- Organization of Panel ERTS 2020. Certification of machine learning for safety critical applications: probable, plausible or impractical?
<https://www.erts2020.org/panels.html>
- Participation: keynote workshop dataIA 2020 Certification of AI-based systems: challenges and promises <http://dataia.eu/ws-safety-ai>

Scientific animation of the theme



Description of the theme agenda

- seminar: every 2 months
- animation meeting: every 2 months

Theme roadmap (year 2, 3 and 4 - cf p38 roadmap)

Year 1: white paper on certifiable AI to be written and completed.