cea list

Characterizing AI Trustworthiness through Formal and Empirical Methods: CEA in PEPR SAIF

Zakaria Chihani

FRANCE 2030

Ce travail a bénéficié d'une aide de l’État gérée par l'Agence Nationale de la Recherche au titre de France 2030 portant la référence « ANR-23-PEIA-0006 »

LSL/AISER role in PEPR SAIF

- Task 1.1: Principled Synthetic Data Generation
- Task 2.1: Open, Modular, Unifying Verification Framework
- Task 2.3: Advanced Neural Network Architectures
- Task 3.3: Generator-Based Properties
- Task 4.1: Monitoring, Harnesses, and Fail-Safe Procedures
- Task 4.2: Principled Training Approaches
- Task 5.1: Verification for Explainability and Explainability for Verification
- Task 5.2: Case-Based Reasoning

Overview of our research

Industry

cea list

Academia

PyRAT

AIMOS

CAISAR

Colibri & co

PARTICUL

Verification	Test	Plateforme	Symbolic	XAI & uncertainty

Overview of our research

Principle: abstract interpretation
- **Conservative** over-approximation of the behaviour of a model
- A property verified on the over-approximation is also verified on **any concrete** behaviour of the model

Target: Neural networks architectures

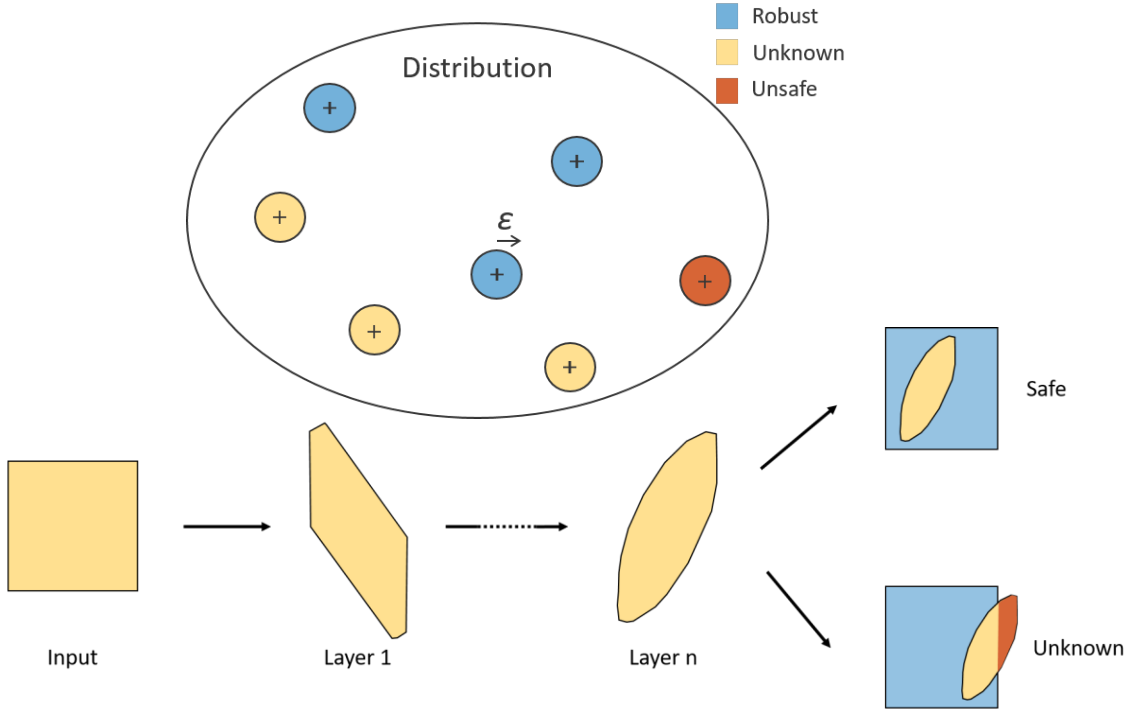
Background:
Decades of use in critical SW and HW verification

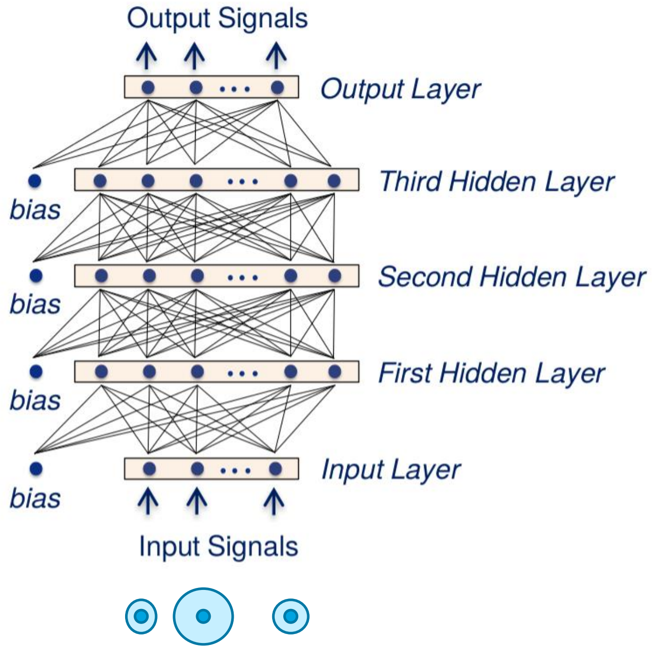
Application:
Verification of functional properties
Verification of robustness to neighbourhood perturbations

PyRAT | AIMOS | CAISAR | Colibri & co | PARTICUL

Verification	Test	Plateforme	Symbolic	XAI & uncertainty
PyRAT	AIMOS	CAISAR	Colibri & co	PARTICUL

Overview of our research





PyRAT | AIMOS | CAISAR | Colibri & co | PARTICUL

Verification	Test	Plateforme	Symbolic	XAI & uncertainty

Overview of our research

PhD in the pipe with LMF, directed by Serge Haddad, on automata and abstract interpretation
PhD ongoing with Inria, co-directed by Caterina Urban, verification, robustification and explainability

Participation in VNN-Comp, industrial applications, academic collaborations (quantized networks - Romania, closed-loop systems with a visitor from Stanford)

Task 2.3: Advanced Neural Network Architectures (RNN, GNN, transformers, quantized networks)
Task 4.2: Principled Training Approaches (certified training and sparsification)
Task 5.1: Verification for Explainability and Explainability for Verification (connections between the two, robustifying explainability with abstract interpretation)

PyRAT	AIMOS	CAISAR	Colibri & co	PARTICUL
Verification	**Test**	**Plateforme**	**Symbolic**	**XAI & uncertainty**

Overview of our research

Principle: Metamorphic testing
- Based on operational domain, describe relations on inputs and the expected relations on outputs.
- Automatically generate a test set to evaluate the satisfaction of these relations.

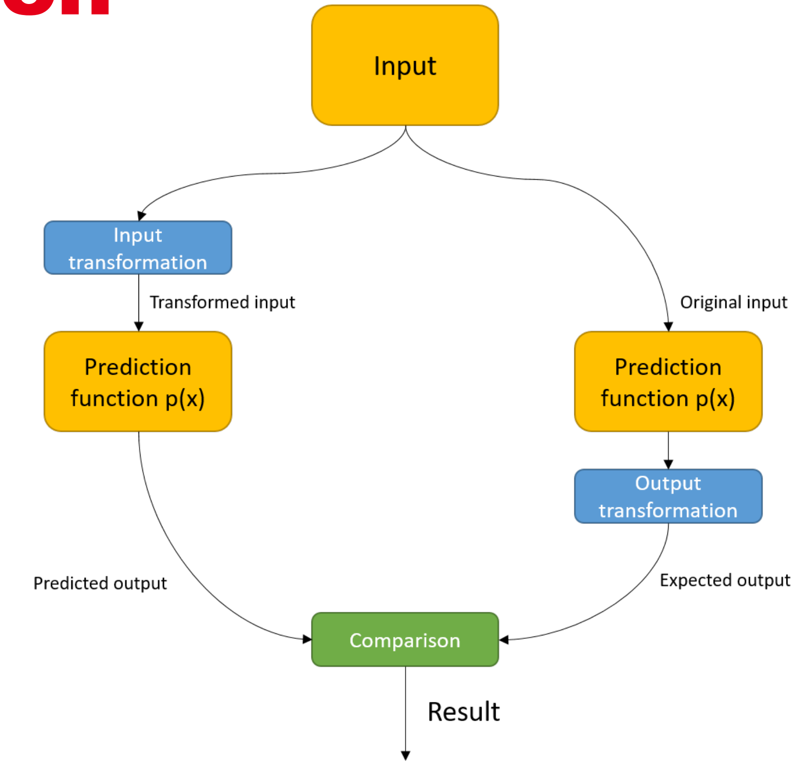
Target: Application and model agnostic: image classification, tabular data, time series, SVM, …
Application: robustness to different luminosity levels, blur, symmetry, ...

Background: Metamorphic testing has been used in software V&V for decades

PyRAT | AIMOS | CAISAR | Colibri & co | PARTICUL

Verification	Test	Plateforme	Symbolic	XAI & uncertainty

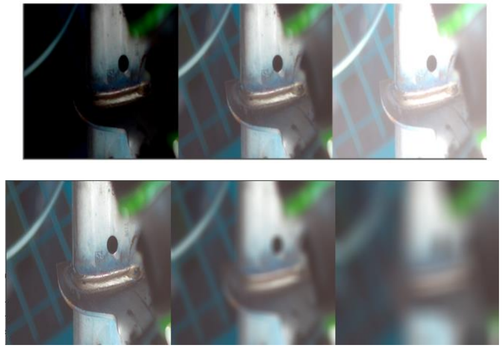
Overview of our research

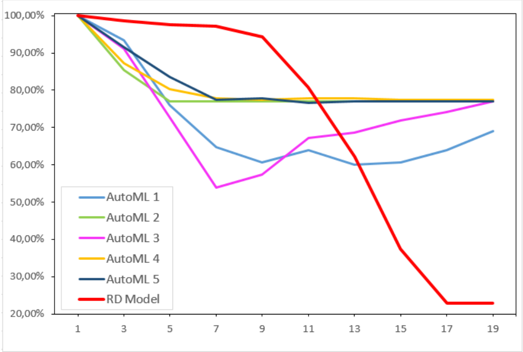


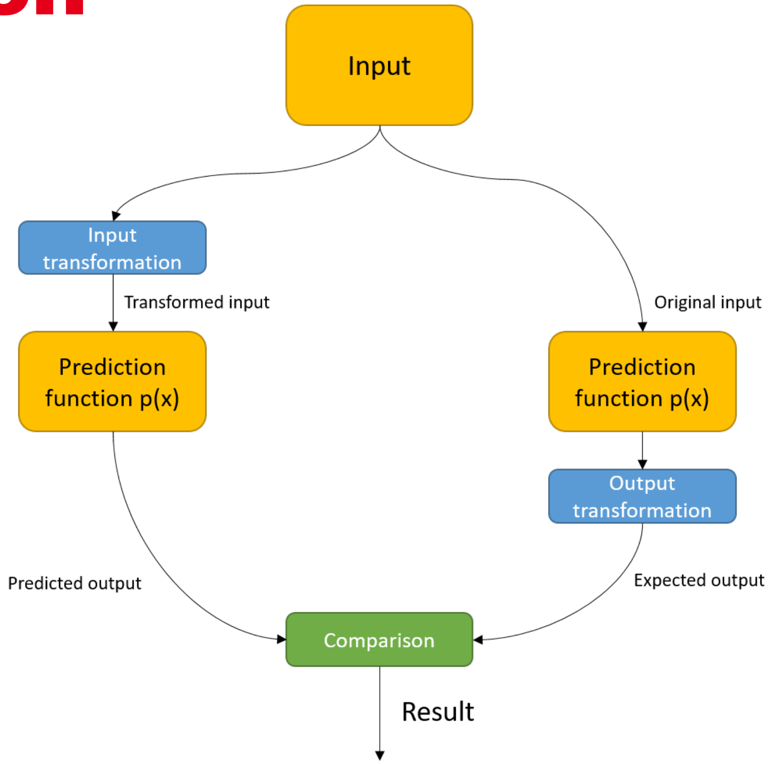
PyRAT | AIMOS | CAISAR | Colibri & co | PARTICUL

Verification | Test | Plateforme | Symbolic | XAI & uncertainty

Overview of our research







PyRAT | AIMOS | CAISAR | Colibri & co | PARTICUL

Verification	Test	Plateforme	Symbolic	XAI & uncertainty

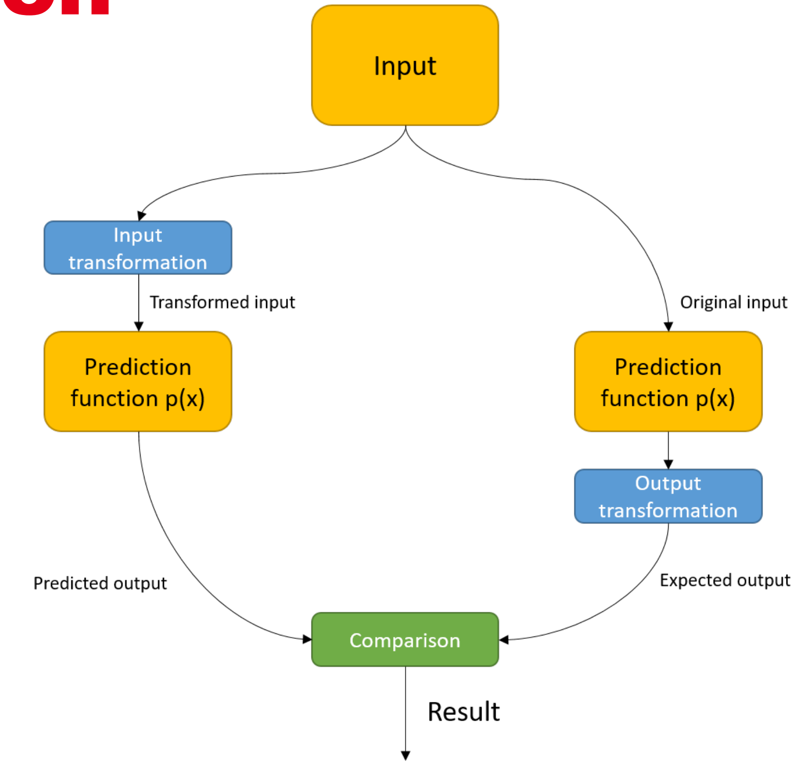
Overview of our research

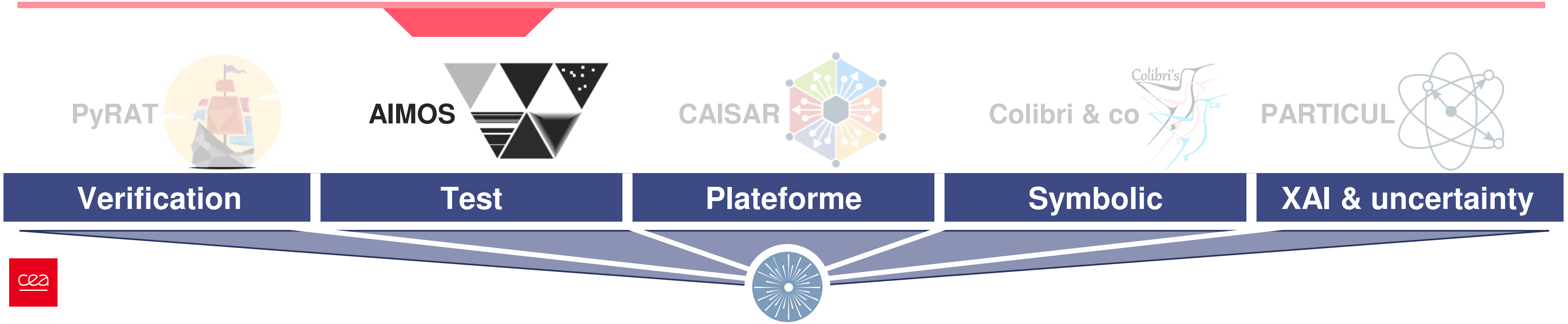
High maturity, GUI in the pipes, few avenues for exploratory research

Available for academic purposes (Lab sessions, courses…)

Contact with Siemens-Germany on using metamorphism to constraint test generation with GAN

Task 1.1: Principled Synthetic Data Generation





Overview of our research

Principle: Maximize coverage of AI models and properties
- Common expressive specification language
- Easy extensibility through clear interfaces
- Heuristic-aided V&V analysis
- Common aggregation of analysis outputs

Target: SVM, Neural Networks, XGBoost models, ensemble models,…

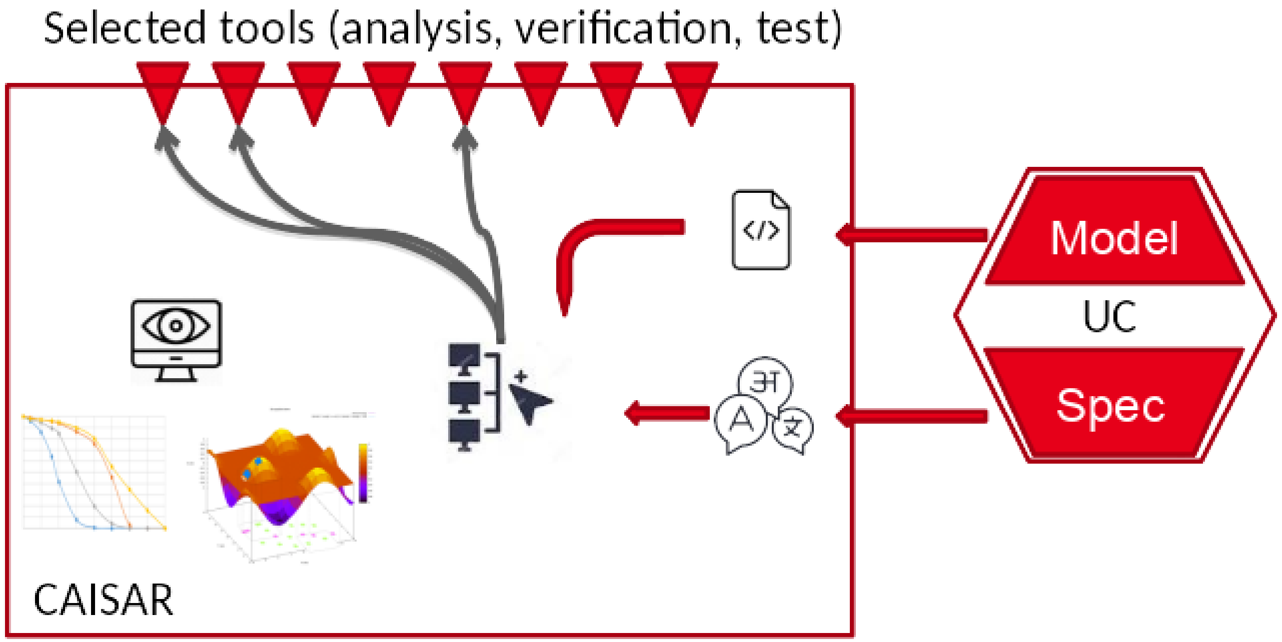
Application: depending on the used plug-ins. Currently includes
- SAVer for SVM
- Colibri for XGboost
- PyRAT, AB-Crown, Nnenum, Marabou for NN

Background: The federative platform strategy for V&V has been successful for critical SW (see, for example, Frama-C and Why3)

PyRAT | AIMOS | CAISAR | Colibri & co | PARTICUL

Verification | Test | Plateforme | Symbolic | XAI & uncertainty

Overview of our research



Overview of our research

On the back-ends : Contacts with various tool providers from academia and private sector

On the platform : Contacts related work team such as Vehicle (Edinburgh, Wales)

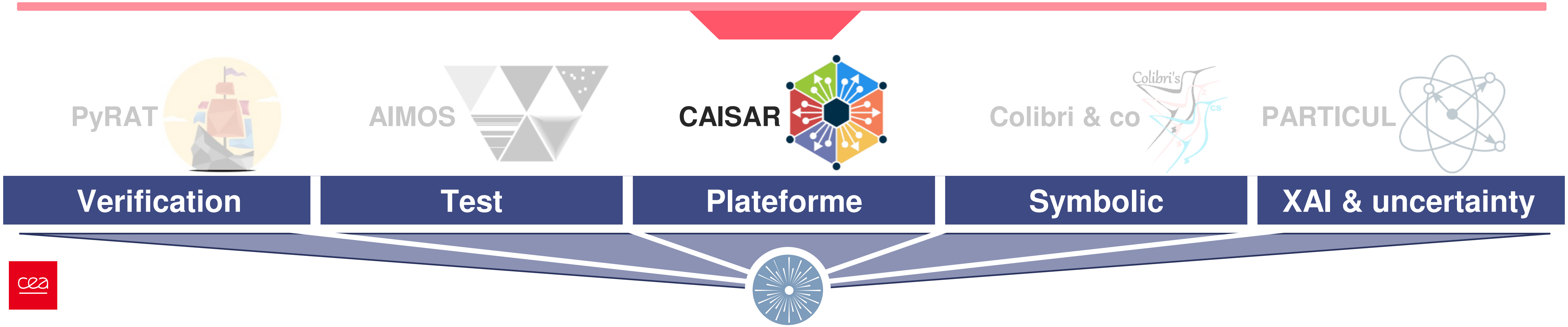
On the usage : neurosymbolic AI (Dortmund), discussion for using it as interface during VNN-Comp

A visitor coming from Sweden.

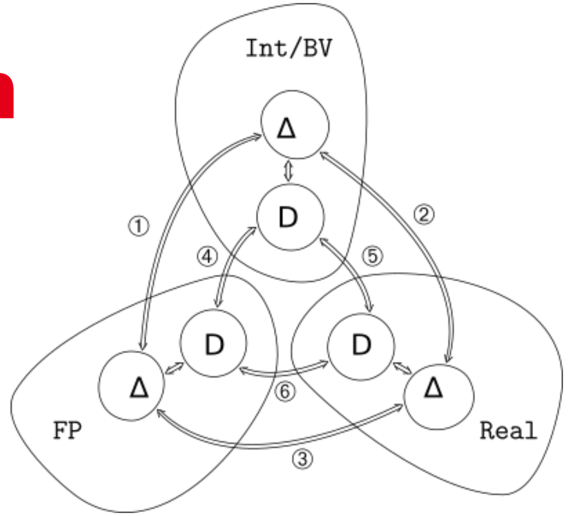
Also available for teaching (lab sessions, courses)

Task 2.1: Open, Modular, Unifying Verification Framework

(don’t worry, we’re not starting from nothing ☺)



Overview of our research



Principle: Safe-by-design Symbolic AI through a constraint solving library

- Separately prove the necessary bricks for constraint solving: Floating-point numbers, integers, bit-vectors, strings, etc.
- Allow for selection of these bricks to tailor the construction of a solver to the needs of the user
- Automatically extract a C implementation of the solver

Target: XGBoost models, embedded software

Application: Energy sector (e.g., IRSN), space (e.g., NASA). Can also be used as a verification tool (winner of SMT-Competition since 2017), which makes it an essential brick of other tools such as Frama-C and GATeL.

Background: Constraint solving is used in several critical software domains

PyRAT | AIMOS | CAISAR | Colibri & co | PARTICUL

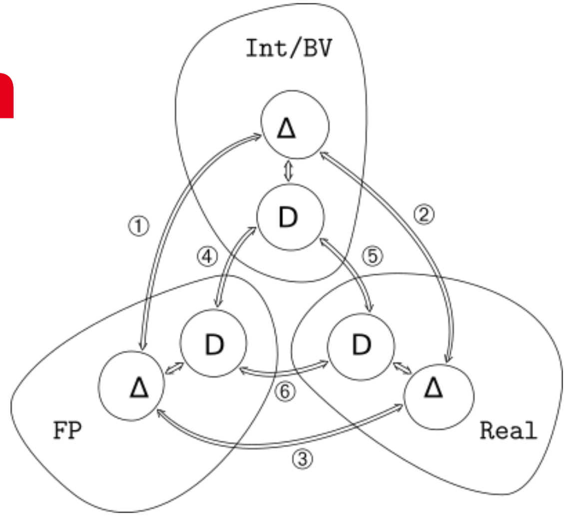
Verification	Test	Plateforme	Symbolic	XAI & uncertainty

Overview of our research

Two derived tools with various purposes
- better connection as a back-end, incorporating (SAT-style) learning,
- verified-by-design : extracted from the proof in Why3

The constraint solving power can play in the enforcing of logical properties

Task 3.3: Generator-Based Properties
Task 4.2: Principled Training Approaches



PyRAT | AIMOS | CAISAR | Colibri & co | PARTICUL

Verification | Test | Plateforme | Symbolic | XAI & uncertainty

Overview of our research

Principle: Detect recurring parts in a dataset through unsupervised learning

- Pluggable: added to an existing backbone, fraction of the size
- Frugal: no need to fine-tune the backbone
- Non-invasive: minimal access to the backbone and a fraction of the data
- Fast: convergence in a few epochs
- Measured: gives confidence measures of the detections

Target: Neural networks

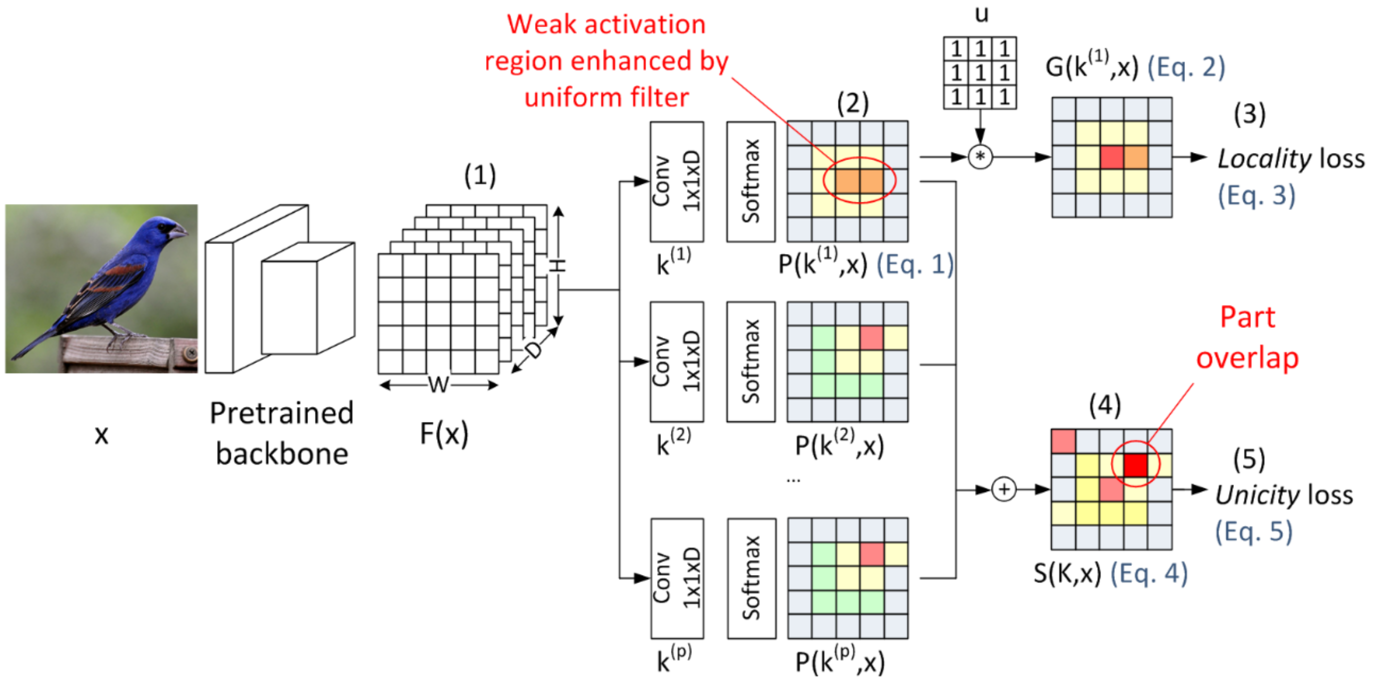
Applications:

- Interpretable out-of-disribution detection
- Boosting classification
- Aided annotation
- Explainability, as a brick of case-based reasoning

PyRAT | AIMOS | CAISAR | Colibri & co | PARTICUL

Verification | **Test** | **Plateforme** | **Symbolic** | **XAI & uncertainty**

Overview of our research



PyRAT | AIMOS | CAISAR | Colibri & co | PARTICUL

Verification	Test	Plateforme	Symbolic	XAI & uncertainty

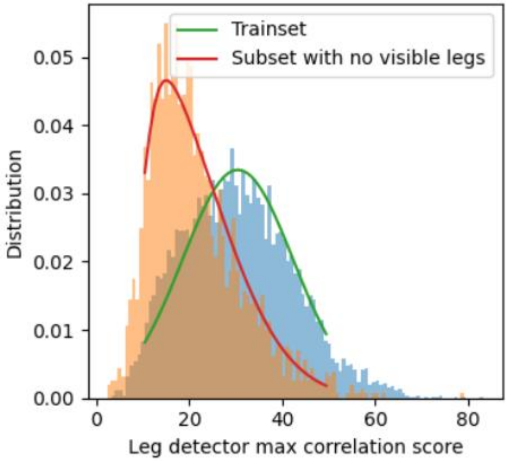
Overview of our research

PARTICUL

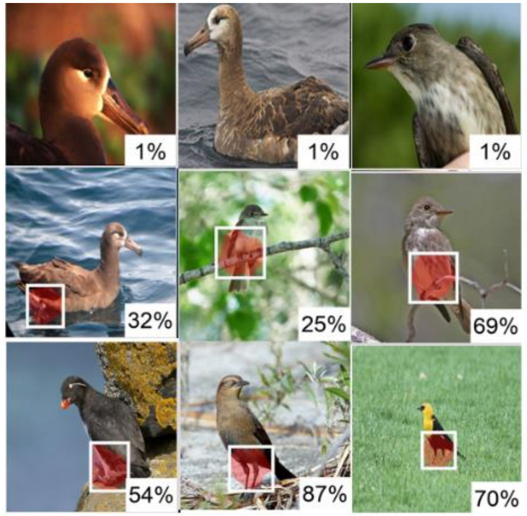
PyRAT | AIMOS | CAISAR | Colibri & co | PARTICUL

Verification	Test	Plateforme	Symbolic	XAI & uncertainty

Overview of our research



(a) Distribution of maximum correlation scores on the CUB-200 training set (in blue) and on a subset containing only images with non-visible legs (red).



(b) Confidence scores and part visualizations on images with non visible legs (top-row) and with visible legs (bottom rows).

PARTICUL

PyRAT | AIMOS | CAISAR | Colibri & co | PARTICUL

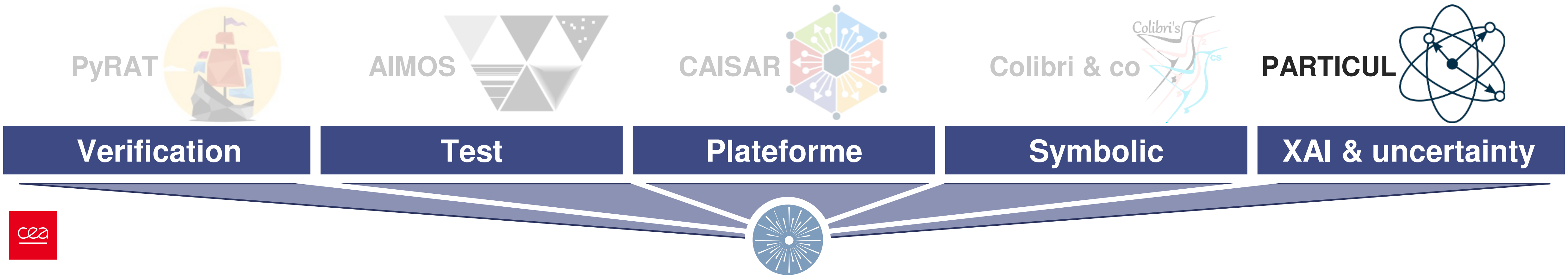
Verification | Test | Plateforme | Symbolic | XAI & uncertainty

Overview of our research

PhD and CDD started on XAI and Formal methods

Work ongoing on a library for protypes (CaBRNet), with national and international contacts (Poland, Netherlands, Bordeaux) – available for (and soon, we hope, used in) XAI course

Task 4.1: Monitoring, Harnesses, and Fail-Safe Procedures
Task 5.1: Verification for Explainability and Explainability for Verification
Task 5.2: Case-Based Reasoning



The people behind the scene

Serge Durand Tristan Le Gall Julien Lehmann Augustin Lemesle Jaouhar Slimi	Augustin Lemesle Aymeric Varasse	Michele Alberti François Bobot Julien Girard Augustin Lemesle Aymeric Varasse	Hichem Ait-el-Hara François Bobot Bernard Botella Bruno Marre	Serge Durand Julien Girard Alban Grastien Jules Soria Romain Xu-Darme
PyRAT	AIMOS	CAISAR	Colibri & co	PARTICUL
Verification	Test	Plateforme	Symbolic	XAI & uncertainty