ULTIMATE

# TrustWorthy AI Cluster

Working towards a Trustworthy AI

#TrustWorthyAICluster

AUTOFAIR SAFEXPLAIN EVENFLOW TALON Tuples ULTIMATE enexa REXASI

On February, the Trustworthy AI Cluster presented and discussed 3 remarkable or promising results and 3 future research/innovation challenges to support building future on-demand Solutions with AI, Data and Robotics.

Here are the insights on the #3 promising results and future innovation challenges chosen by Trustworthy AI Cluster partners.


## #1 Considering hybrid AI to build or increase trustworthiness

In comparison with pure ML/DL applications, hybrid AI (especially neuromorphic approaches) enable to take into account symbolic bases (e.g. temporal properties, rules expressed by humans) to complete their evaluation, V&V using robust methods (like formal ones).

## #2 Building trustworthy AI activities along the whole AI life cycle

Legacy AI-based systems but also innovative ones have to consider an end-to-end approach to design, increase, maintain trustworthiness at different engineering levels, from algorithm, System / SW /HW level, with consistency.

## #3 End-to-end Trustworthy AI including Value Sensitive Design based

Trustworthy AI (even if declined from ethical principles) is mainly driven by quantifying technical criteria. The proposed experimental approach consists in considering the ethical dimension, at the same level than the other technical trustworthy AI criteria, through some questionnaires filled by different stakeholders (e.g. customer, AI developer, AI V&V). It may impact the way that they think, design, implement, V&V the AI solution.

## Scalability and transferability of trustworthy frameworks

Create a layered approach to trustworthiness (data, control, human layers).

Explain how the framework can scale with future and more demanding applications. Also consider and propose ways to transfer/adapt the trustworthiness between different verticals

Newsletter #6 2025

Lukasiewicz PIAP LNE tecnalia iTTi Robotnik THALES CBRNE Ltd ThalesAlenia Space KTH

(i.e., from robotics to manufacturing).

## AI-based risk management environment

Build a dedicated environment to manage AI risks (potentially based on the NIST AI RMF) for a specific critical domain based on an AI risk repository (potentially inspired from the MIT one). This environment should be flexible enough to instantiate the AI risks and management rules applicable to a given critical domain. It should fulfil the requirements of AI regulation document(s) depending on the target risk level to prevent or mitigate risks according the chosen policy. It could be coupled with classic risk management tooled-up methodologies (e.g. STPA).

## Collecting and organizing good practices

Colect the returns of Experience, including assumptions in terms of trustworthiness to make the AI life cycle more mature.

Find out the presentations here:
https://adr-association.eu/events/future-ready-demand-solutions-ai-data-and-robotics

### Interview with Dr. Delphine Longuet - cortAIx Labs formal verification expert

*In the Ultimate project, my job was to demonstrate mathematically, using formal verification tools, the robustness of anomaly detection to certain signal perturbations.*
*Seeking to develop hybrid AI algorithms that are reliable, explainable and robust algorithms is the aim of Trustworthy Hybrid AI.*
*On an algorithm that detects different types of anomalies in time series coming from satellites sensors, we can now prove that, for each time series in the training set, that, if a low-level noise disturbs the signal, the algorithm always detects the right anomaly.*
*This is the first time we have been able to apply this type of verification to an industrial case study, the one of Thales Alenia Space.*

### Event

On the 7th of April, the TALON Project, one of our TrustworthyAI cluster partners, presented a webinar on AI you can trust. Trustworthiness versus existing standards and additional AI risks and challenges towards standardization were explained as in the Faith process.
Find out more here: https://talon-project.eu/


ULTIMATE

This newsletter is not formatted for paper printing. Preserve the environment and only print this page if necessary.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.