

ATRASS#1, Jan 23, 2025

The first edition will start with a presentation of the supporting programmes. Future sessions will mainly contain scientific talks and discussions.

Agenda:

4:00pm CET Introduction of the supporting programmes (10 minutes each):

(by alphabetical order)

CERTAIN by DFKI

Confiance.ai by IRT SystemX

IVADO/Confiance.ia qc by Polytechnique Montreal

Responsible AI UK by U Southampton

UWF programme by Univ. West Florida

Zertifizierte KI by Fraunhofer IAIS

5:00pm CET Scientific seminar by Anita Prinzie, Datashift

Towards a deep integration of Trustworthy AI and AI Risk Management for all AI systems

Abstract: Trustworthy AI and AI Risk Management share the common goal of ensuring that AI systems can be trusted, yet they approach this objective through distinct but compatible methods. This presentation explores the core principles of Trustworthy AI and AI Risk Management, highlighting the complementarity of these two domains, which have largely developed independently. We argue for the benefits of deeply

integrating Trustworthy AI and AI Risk Management across all AI systems, not limited to high-risk applications mandated by the AI Act. Finally, we outline a vision for achieving this integration, addressing the scientific challenges that must be overcome to create a unified framework.