

Trusted AI in the age of generative AI and complex AI systems

November 9th, 2023

Rodica Mihai, NORCE Energy

André Meyer-Vitali, DFKI

Emmanuel Kahembwe, VDE

ADRForum – Versailles, 2023

Agenda



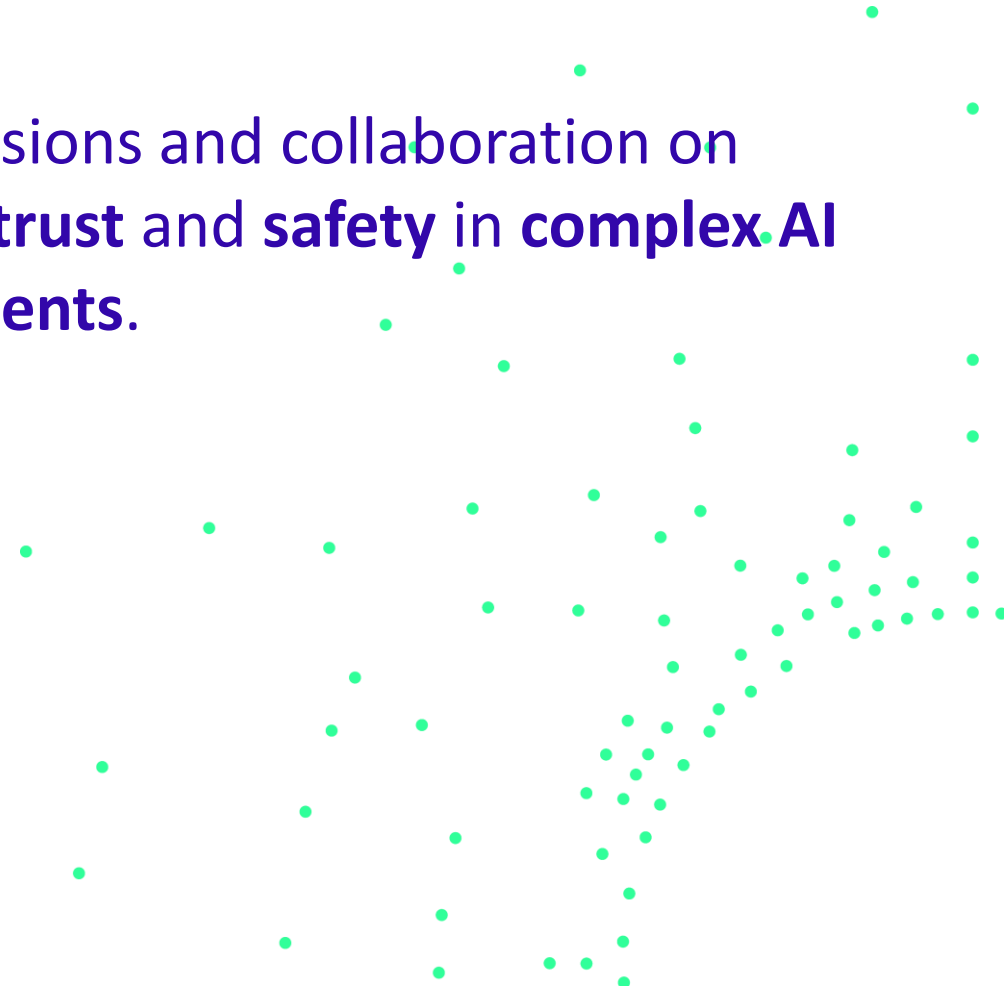
- Short introduction, Rodica Mihai
- Principles of Trusted AI, Dr. André Meyer-Vitali
- Trust in the Age of Generative AI, Dr. Emmanuel Kahembwe
- Trust and safety in the context of complex AI systems: challenges related to verification and validation, Dr. Rodica Mihai
- Discussion session



Short introduction



- What is trusted AI and how can we achieve it?
- Main objective of the workshop is to foster discussions and collaboration on mapping challenges and opportunities related to **trust** and **safety** in **complex AI systems**, with an emphasis on **high-risk environments**.



Short introduction



- Some background:
 - A background of trusted and trustworthy AI will be given, with some of the key aspects of trusted AI systems being presented.
 - How to guarantee ethical characteristics of AI systems and how to design them according to system engineering principles.
 - We will exemplify some of the challenges we are currently facing in the context of generative AI and increasing complexity of AI systems, in a cross-domain perspective.



Short introduction



- Aim to:
 - trigger discussions on challenges and current trends on trusted AI.
 - discuss and map current gaps.
 - share examples.

Trust and Safety

Complex AI systems

High-risk AI

