13th Sept 2023 9:00 - 17:30 CEST

TRUSTED AI
THE FUTURE OF CREATING ETHICAL AND RESPONSIBLE AI SYSTEMS

Theme Development Workshop

Identify common goals between academia & other relevant stakeholders, and define promising approaches for European research and innovation for Trustworthy AI.

The workshop will be held online via Zoom with a mixed programme of presentations and in-depth discussions about specific sub-topics in smaller groups (Breakout sessions). This gives the participants the opportunity to discuss with selected experts and contribute to the strategic research and innovation agenda for AI in Europe.

Workshop programme

09:00 – 09:15	**Welcome & Objectives**
09:15 – 09:35	**Principles of Trusted AI** ***André Meyer-Vitali, DFKI***
09:35 – 09:50	**Role of the EU and orientation of EU policy making in relation to trustworthy, responsible and ethical AI** ***Antoine Alexandre André, DG CNECT***
09:50 – 10:00	**Coffee Break & Socialising**
10:00 – 11:30	**Parallel Breakout sessions**
11:30 – 12:30	**Plenary presentation of key findings from the Breakout sessions**
12:30 – 13:30	**Lunch break & Socialising**
13:30 – 13:45	**Ethical AI** ***Meeri Haatja, Saidot***
13:45 – 14:00	**Responsible AI in the industry** ***tba***
14:00 – 15:30	**Parallel Breakout sessions**
15:30 – 15:45	**Coffee Break & Socialising**
15:45 – 16:45	**Plenary presentation of key findings from the Breakout sessions**
16:45 – 17:30	**Closing & Socialising**

Please register here.

We invite the community to suggest further topics of interest for the breakout sessions. Please use the online application form for your suggestions.

Breakout sessions

Breakout session 1:
AI explainability for vision tasks
This session will discuss the present and future of AI explainability for visual data classifiers and other vision tasks, how explanations can be presented to the users, and what we can expect to understand from these explanations.

Breakout session 2:
Ethical considerations and new challenges of Generative AI
This session aims to explore the risks and challenges raised by generative AI from an interdisciplinary perspective (legal, ethical, societal, technical, cybersecurity).

Breakout session 3:
Rigorous vs empirical AI privacy: Where is the middle ground for defining and evaluating privacy in complex algorithms?
This session will discuss the relevance of epsilon as a definitive measure of privacy loss in the context of complex algorithms implementing differential privacy and the proliferation of empirical measurements of privacy via attacks.

Breakout session 4:
Monitoring progress in interpretable AI
It is well studied how to measure the accuracy of machine learning predictors; it is less trivial to monitor progress in developing models interpretable by humans. We propose to bring together an interdisciplinary group of experts (legal, regulational, technical aspects) to outline the requirements for such monitoring and possible ways to approach this problem.

Breakout session 5:
Causality and Trust
Causal models can improve the trustworthiness of AI systems (Causality for Trust, C4T). Besides precision and accuracy, which are fundamental to trustworthiness in AI, they are transparent, reproducible, fair, robust, privacy-aware, safe and accountable.

Breakout session 6:
Robustness/Verification
This session looks at technologies to strengthen the secure use of AI technologies. There will be a discussion on certifiable robustness, resilience and recovery, and uncertainty and safety in decision making. Finally, the relationships between robustness and privacy, explainability, and fairness, to rule out potential trade offs or define suitable mitigation strategies will be discussed.

Breakout session 7:
AI/ML Benchmarking
This session reflects on ways to assess technologies and methodologies in real-world conditions. While looking at a set of examples, the process of measuring the maturity level and impact of an application will be demonstrated. Appropriate ways to identify innovation early on, protect it and finally transfer generated knowledge to industry and society will be discussed.

Breakout session 8:
AI Ethics: from principles to practice. Putting "ethical" and "responsible" AI into action
The session will focus on operationalizing the AI ethical guidelines and principles. It will reflect on the shifting approach from high-level ethical principles towards legally binding obligations (e.g. in the AI Act) and practical tools (e.g. the Human Rights Impact Assessments).

Breakout session 9:
Meaningful Human Shared Control
Over the past years, we have seen a number of guidelines promoting 'human-in/on/out of-the-loop' approaches to ensure human control and oversight over AI systems. This topic explores the interplay between the dynamic transfer of tasks and ensuring the long-term control over the socio-technical system.

Breakout session 10:
Human Oversight and Explainability for AI
This session will look at architectures, mechanisms and methods capable of generating meaningful and evidence based assurance which is necessary to secure and maintain the safety and security dimensions of AI systems.

Breakout session 11:
Trusting Each Other
For collaborative decision-making (CDM), it is essential that each human and agent is aware of each others' points of view and understands that others possess mental states that might differ from one's own - which is known as a Theory of Mind (ToM).

Breakout session 12:
Human-Aligned Video AI
Video-AI holds the promise to explore what is unreachable, monitor what is imperceivable and to protect what is most valuable. But what exactly defines human-aligned video-AI, how can it be made computable, and what determines its societal acceptance?

Breakout session 13:
Trustworthiness in Robotics: at home, at work, and in the city
How to ensure trust of humans to a robot? What are the hindrances to build that trust, at home, at work, in the city? How to ensure trust with people suffering from cognitive, physical or sensorial deficiencies? The session will take inspiration from the guidelines of the High-Level Expert Group on AI which recommends that AI systems should meet a set of requirements to be deemed trustworthy.

Breakout session 14:
Ethics in Games AI
Games is an application domain of AI research that is often overlooked when discussing responsible AI. This session aims to challenge this by discussing the unique challenges that appear in the games environment (E.g. need for believable characters) while also satisfying ethical values.