CHALLENGE: TRUST & TRUSTWORTHINESS

ROEL DOBBE & MYRTHE TIELMAN
R.I.J.DOBBE@TUDELFT.NL – M.L.TIELMAN@TUDELFT.NL

What are we about?

In this challenge, we aim to inform the creation of trustworthy HI, and build and calibrate trust in human-AI systems.

- Our mission makes room for work at three key levels.
  1. Trustworthiness of technical HI advances
  2. Trustworthiness of human-AI interactions and the relation to the calibration of appropriate trust
  3. Trustworthiness of the broader sociotechnical system in which HI operates

What are our plans?

The following questions will be further refined in conversation with the other members:

- RQ1: What do trustworthiness and trust mean at the level of HI technology, HI interactions and sociotechnical HI systems?
- RQ2: How do we evaluate trustworthiness and trust, both at each level and across the different levels?
- RQ3: How can we design for trustworthiness in HI systems and create HI which is trusted appropriately?

We start with aligning perspectives and sharing insights

- First meeting new series: Sept 9
- 4-6Weekly meetings
- 1 presentation of how project aligns with challenge (30min)
- Discussion of 1 paper aligned with the challenge (30min)