Trustworthy AI regulations and their industrial/societal implications
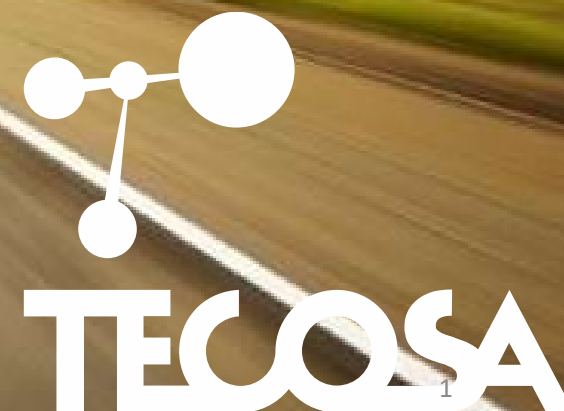
Martin Törngren and Rafia Inam, Mechatronics,
KTH Royal Institute of Technology, Stockholm

digital futures

TECOSA

Agenda

13:00 – Session 1: Keynote and Invited short talks (1h+30 mins)

• Keynote talk by Tatjana Evas (EC) (30 mins, 20+10)

• Short talks to set the context for the industrial panel (10+5)!
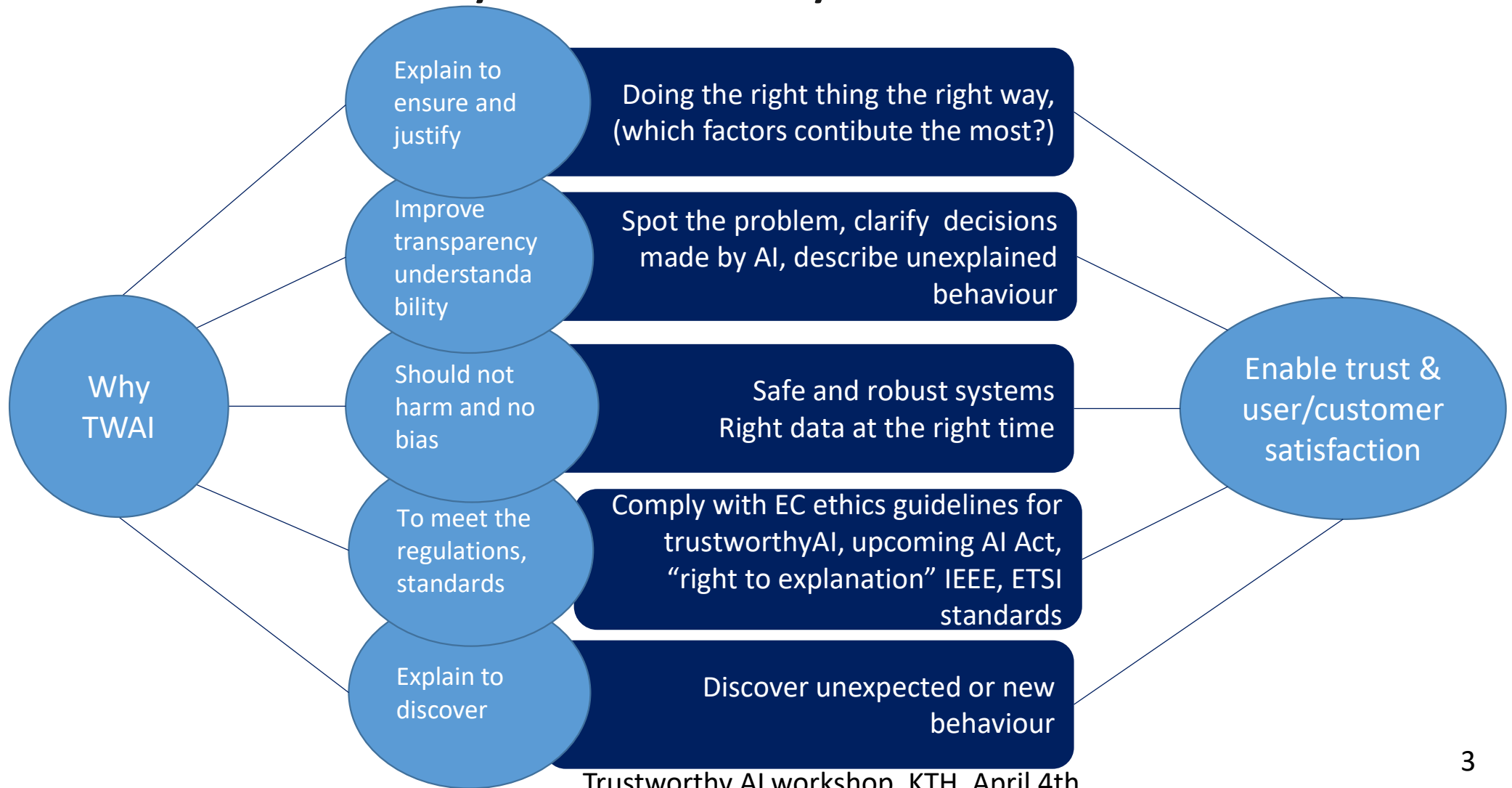
14:35 – change to break out groups - Leg stretch, coffee (10 mins)

14:45 – Session 2: Break out groups and brief reporting

15:25 – Session 3: Panel debate – industrial responses (40 mins)

16:05 – Next steps and Wrap up (10 mins)

Trustworthy AI for Society

Why TWAI

- **Explain to ensure and justify** — Doing the right thing the right way, (which factors contibute the most?)
- **Improve transparency understandability** — Spot the problem, clarify decisions made by AI, describe unexplained behaviour
- **Should not harm and no bias** — Safe and robust systems Right data at the right time
- **To meet the regulations, standards** — Comply with EC ethics guidelines for trustworthyAI, upcoming AI Act, "right to explanation" IEEE, ETSI standards
- **Explain to discover** — Discover unexpected or new behaviour

Enable trust & user/customer satisfaction

Trustworthy AI workshop, KTH, April 4th

3

Trustworthy AI Regulations

Development of regulations around AI is progressing rapidly faster than any other recent technological advance

Definition
— *Trustworthy AI* is a term used to describe "AI that is lawful, ethically adherent, and technically robust" [EC Ethics Guidelines].
— In simple terms: perform as intended and do no harm

Regulations Act
— EC Ethics Guidelines, assessment list for trustworthy AI, EU Artificial Intelligence Act
— Google, Microsoft, Amazon, ... Work on Responsible AI
— Recent published report U.S. National Security Commission for Artificial Intelligence (NSCAI)
— Blueprint for an AI Bill of Rights - The White House Released by the U.S. Office of Science & Technology Policy– Oct 2022, US plan to promote ethical AI- The Biden administration is introducing new funding and policy guidance for developing artificial intelligence, May 4th 2023
— U.S. Executive Order on the safe, secure and trustworthy development of AI, Nov 2023
— Canadian Artificial Intelligence and Data Act.
— Stanford Center for Human-Centered AI: https://hai.stanford.edu/research/ai-index-2022
— IEEE, ETSI standards, …

Trustworthy AI Guidelines by

Human Agency and Oversight	**Transparency**	**Privacy and Data Governance**	**Accountability**
Human autonomy is paramount	Usage transparency, explain outcomes	Respects data privacy and user consent	Traceable and auditable

Diversity, Non-discrimination & Fairness	**Technical Robustness and Safety**	**Societal & Environmental wellbeing**
Unbiased and non-discriminatory	Resilient to attacks and uncertainty	Respect human rights and environmental wellbeing

EC Assessment List for Trustworthy Artificial Intelligence (ALTAI)

Workshop overall motivation and goals

Improve awareness of ongoing regulatory changes!

Promote dialogues and interactions between different stakeholders and communities, especially decision makers, legal, research and industrial practitioners!

Discuss impact of regulation on different domains/industries

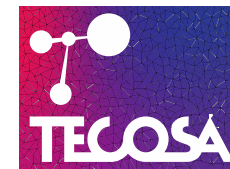Increase networking possibilities among different communities

RI0

Regulations/ Standardization

Academia

Industry

Trustworthy AI

Research

Business / IPR

Society

Agenda

13:00 – Session 1: Keynote and Invited short talks (1h+30 mins)

- Keynote talk by Tatjana Evas (EC) (30 mins, 20+10)
- Short talks to set the context for the workshop! (10+5)
    - Fredrik Heintz, Professor LIU (AI)
    - Eduardo Gill-Pedro, Associate professor, Lund University (faculty of law)
    - Anna Felländer, Anch AI
    - Elisabeth Thand Ringqvist, Member of Parliament, Vice Chair Committee on Industry and Trade
    - Hans Hedin, Swedish Post & Telecom Authority (PTS)

14:30 – Leg stretch, coffee, and change to break out groups (15 mins)

14:45 – Session 2: Break out groups and brief reporting

- 20 min. work session in groups, (chair/notetaker per group to provide a summary with 2 highlights)
    - Main problems to be resolved? Most important action ahead?
- 20 min: Brief summaries per group (1 slide per group)

15:25 – Session 3: Panel debate – industrial responses (40 mins)

- Shiva Sander-Tavallaey (ABB)
- Maria Ramstedt (Ericsson)
- Luis Martinez (Volvo cars)
- Anna Sööder (Schibsted)

16:05 – Next steps and Wrap up (10 mins)

16:15 – *End of Day*