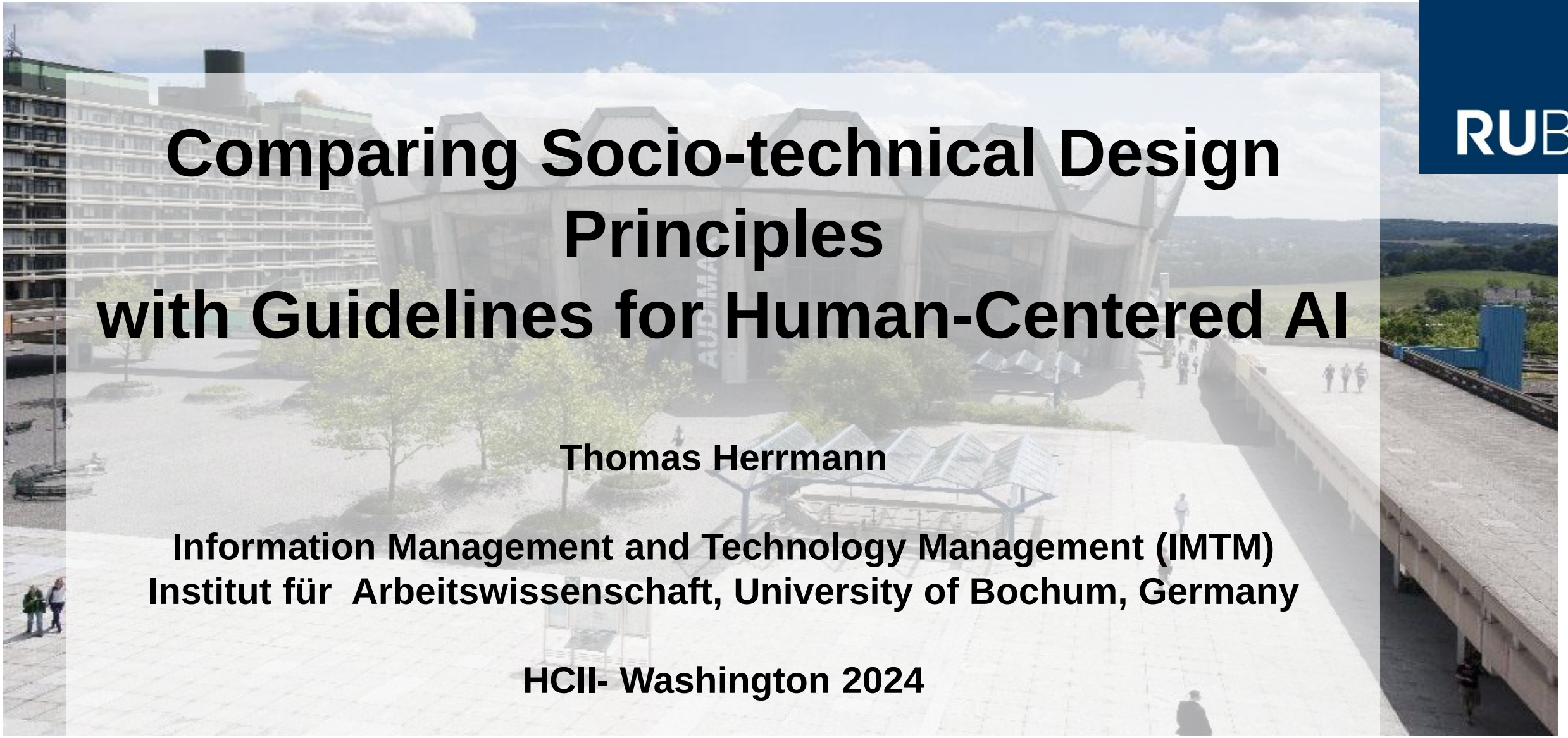


Comparing Socio-technical Design Principles with Guidelines for Human-Centered AI

Thomas Herrmann

Information Management and Technology Management (IMTM)
Institut für Arbeitswissenschaft, University of Bochum, Germany

HCII- Washington 2024

IAW
INSTITUT FÜR ARBEITSWISSENSCHAFT

IMTM

Background

Background 1 – Human Centered Artificial Intelligence (HCAI) and Socio-Technical Systems

"An **Interactive Human Centered Artificial Intelligence** is an Artificial Intelligence that enables **interactive exploration and manipulation** in real time and is designed with a clear purpose for human benefit while being transparent about who has control over data and algorithms. (Schmidt, 2020)"

aiming at using the **complementary strengths of human intelligence and AI** to behave more intelligently than each of the two could be in separation and where

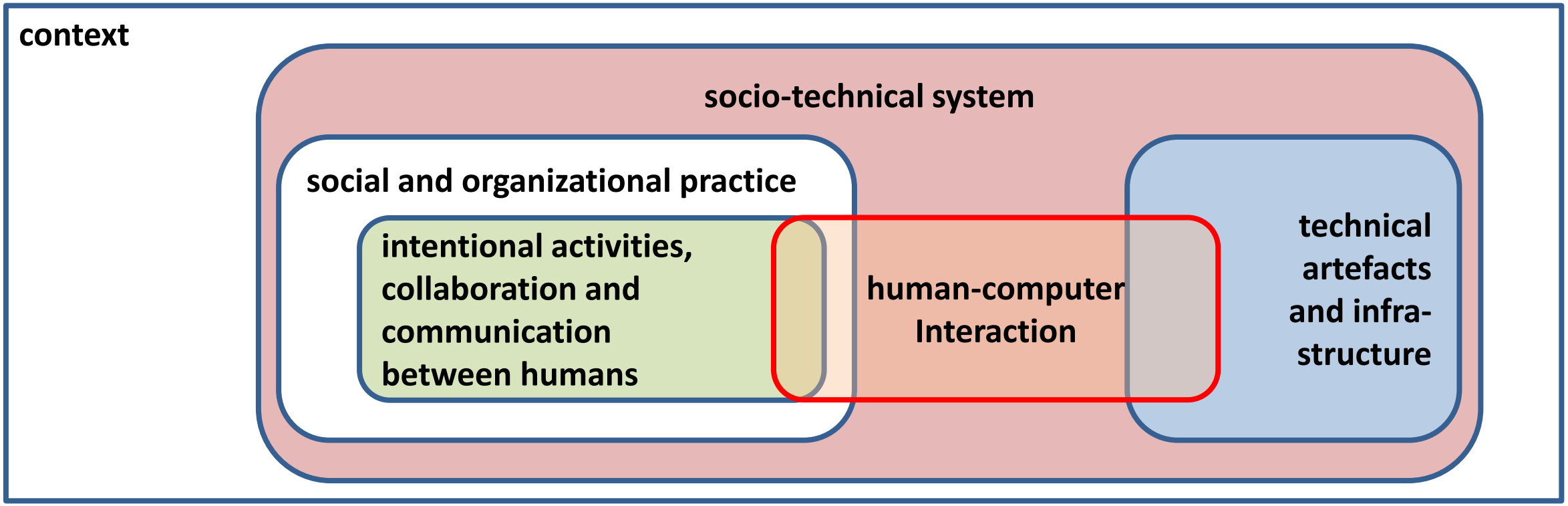
socio-technical ensembles

and its human and AI parts can co-evolve to improve over time,

(Dellermann et al., 2019)

Background 2 – Socio-technical Systems

Socio-technical system as systematical intertwinement of social and organizational practices with technical artefacts and infrastructure
HCI as an enabler of this intertwinement, beside other measures, such as training etc.



Background 3 – Principles to support Socio-technical Design

Socio-technical design is not only about designing technology that includes a socio-organizational requirements,

but also about developing social and organizational practices that complement technical functionality and helps overcome technical shortcomings.

Cherns (1976, 1987)
1. Compatibility with objectives
2. Minimal critical specification of rules
3. Variance control
4. Boundary location of interdependent roles
5. Information flow
6. Power and authority for access to resources
7. Multifuncionality
8. Support of congruence
9. Transitional organization
10. Incompleteness

Mumford (1983) – STS should support following fits
1. Knowledge Fit
2. Psychological Fit
3. Efficiency Fit
4. Task Structure Fit
5. Ethical Fit (social values)

Background 3 – Principles to support Socio-technical Design

Cherns (1976, 1987)
1. Compatibility with objectives
2. Minimal critical specification of rules
3. Variance control
4. Boundary location of interdependent roles
5. Information flow
6. Power and authority for access to resources
7. Multifuncionality
8. Support of congruence
9. Transitional organization
10. Incompleteness

Mumford (1983) – STS should support the following fits
1. Knowledge Fit
2. Psychological Fit
3. Efficiency Fit
4. Task Structure Fit
5. Ethical Fit (social values)

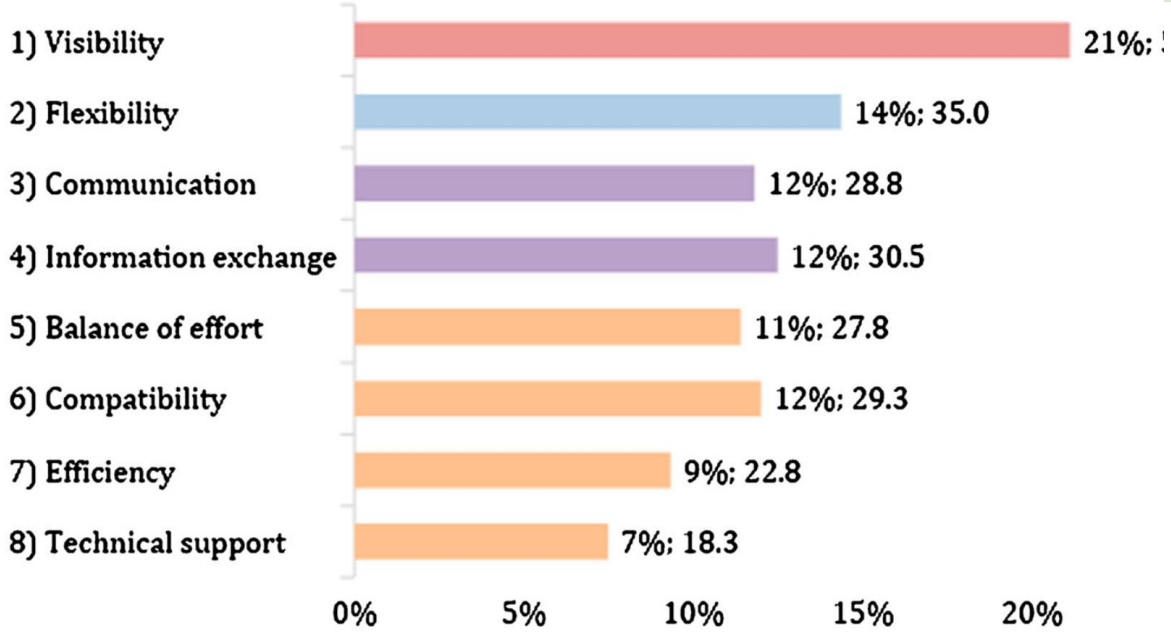
Short comings: new technical developments and new research areas are not considered

Background 4 – Revised Principles on an empirical and interdisciplinary basis

Empirical fields of the problem bases
1. Smart factories
2. Health care
3. Learning support
4. Creativity support

Research areas
1. Usability
2. Computer supported Cooperation
3. Privacy
4. Process Management
5. Job Design

306 Problems



Grouping of 173 principles, guidelines, …

Background 5 – Extended Principles to support Socio-technical design

Cherns (1976, 1987)
1. Compatibility with objectives
2. Minimal critical specification of rules
3. Variance control
4. Boundary location of interdependent roles
5. Information flow
6. Power and authority for access to resources
7. Multifuncionality
8. Support of congruence
9. Transitional organization
10. Incompleteness

Mumford (1983) – STS should support following fits
1. Knowledge Fit
2. Psychological Fit
3. Efficiency Fit
4. Task Structure Fit
5. Ethical Fit (social values)

Herrmann et al. 2021 - Heu tics
1. **Visibility** abou ask ha ng and fee dback abo s
2. **Fle ty** for ha ing lead stem
3. **Com**
4. ck
ced ing of
6. uirements, petencies and the sys res
7. **Effi cy**-oriented allocation of tasks for ursuing holistic goals
8. **Supportive technology** and resources for productive and flawless work

Short comings: AI applications are hardly considered

Background 6 – AI-oriented principles

SLR, (18 papers out of 795 related to HCAI and guidelines or principles)
1. **Transparency,** (Awareness, Comprehensibility, explainability, explorability, traceability)
2. **Autonomy** (human agency, oversight, freedom, human in the loop, controllability)
3. **Accountability** (Responsibility, trust)
4. **Benefits and well-being** (sustainability, minimizing stress, anxiety, frustration)
5. **Fairness** (responsibility, justice, human values, dignity, solidarity, diversity, inclusiveness)
6. **Privacy** (Secrecy, protection of personality,data governance, limited reachability)
7. **Variance** (Variability, Imperfection)
8. **Safety** (robust, reliable, accurate, no harm integrity)

Background 7 – AI-oriented and socio-technical principles

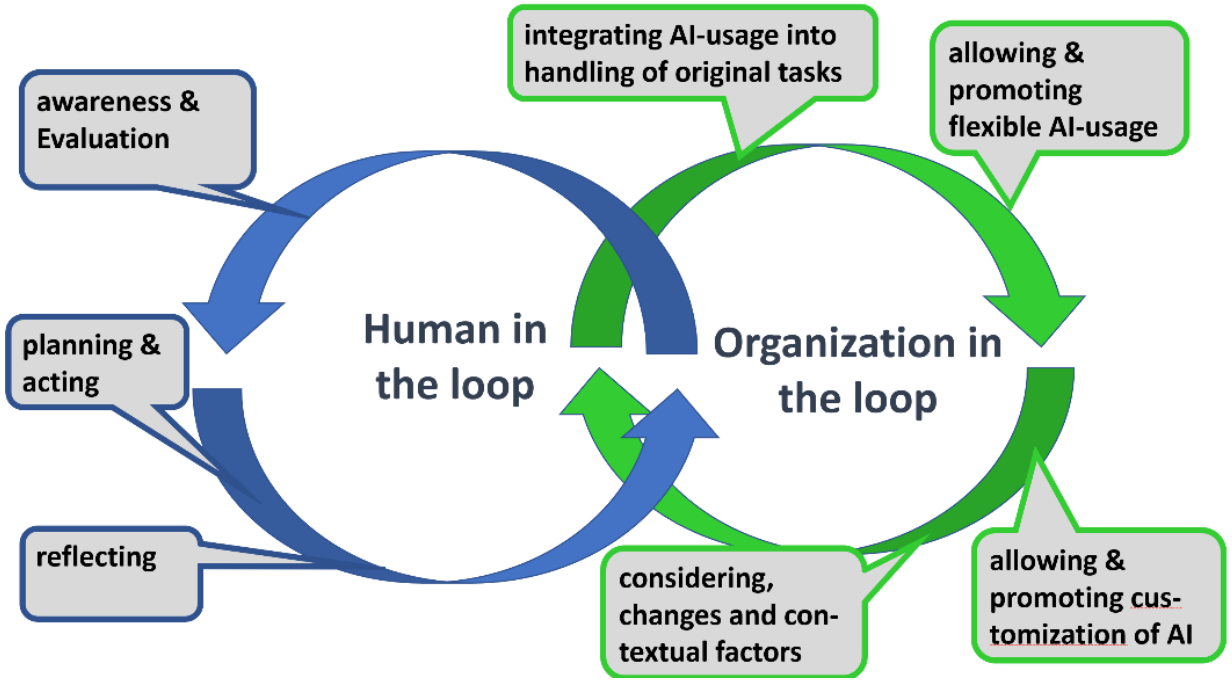
SLR, (18 paper out of 795related to HCAI and guidelines or principles)
1. **Transparency,** (Awareness, Comprehensibility, explainability, explorability, traceability)
2. **Autonomy** (human agency, oversight, freedom, human in the loop, controllability)
3. **Accountability** (Responsibility, trust)
4. **Benefits and well-being** (sustainability, minimizing stress, anxiety, frustration)
5. **Fairness** (responsibility, justice, human values, dignity, solidarity, diversity, inclusiveness)
6. **Privacy** (Secrecy, protection of personality,data governance, limited reachability)
7. **Variance** (Variability, Imperfection)
8. **Safety** (robust, reliable, accurate, no harm integrity)

Socio-technical Heuristics	Related AI aspects
1. **Visibility**	Transparency, Accountability, Privacy
2. **Flexibility and evolution**	Autonomy, Variance
3. **Communication support**	Benefits and wellbeing, accountability, selected aspects of privacy: limited reachability,
4. **Information exchange**	Privacy and data governance, selected aspects of safety: data quality, integrity and access;
5. **Balance of effort and benefit**	Benefits and wellbeing, selected aspects of fairness, such as promotion of human values
6. **Compatibility**	Selected aspects of fairness: avoidance of biases
7. **Efficiency**	Benefits and wellbeing
8. **Supportive technology**	Safety

Findings

Result 1 – Keeping the organization in the loop

In a *human-in-the-loop* (HITL) system, a human operator is a crucial component of an automated control process, **handling challenging tasks of supervision, exception control, continuous improvement.**
Socio-technical perspective: Technically supported **HITL only works if management decisions and organizational practices** authorize, prepare and promote the human activities!

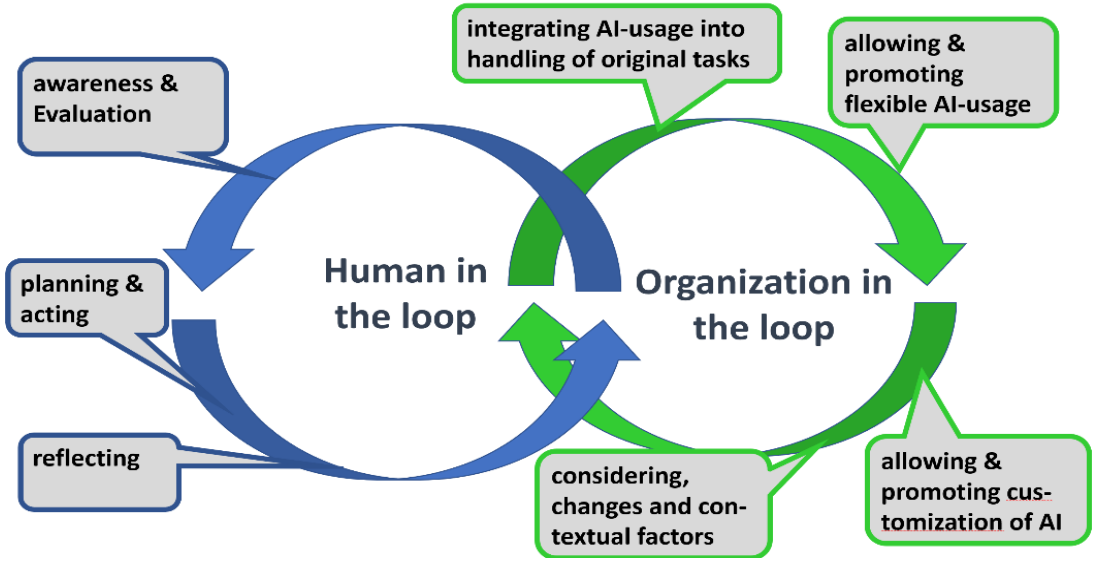


Example:
HCI for controllability and oversight is not sufficient – the management must allow, prepare and promote workers to take over control.
Functionality for explainability is not sufficient – people must have time to use it and support to understand explanations.

Result 2 – Organizational practices can help to overcome shortcomings of AI

Instead of absolute technical reliance and robustness,
compensation by organizational measures is relevant

Example:
Incomplete explanations can be completed by human experts
Hard to understand explanations can be translated by human experts.



Result 3 – AI can help meet socio-technical principles

STS principles or heuristics do not only help design AI applications

AI can also help to realize the principles

Examples:

AI can

- ***serve as a gatekeeper to help regulate reachability in the privacy context***
- ***help to trace the origin of information***
- ***help optimize efficiency***
- **support finding experts who can help with explaining AI**

Result 4 – Autonomy and oversight must be extended with possibilities for mutual evolution

Socio-technical systems are basically characterized as being a subject of continuous transformation
→ This characteristic is also relevant for AI

Examples:

- ***Interactive Machine Learning***
- ***Continuous adjustment to contextual changes***
- ***Continuous exploitation of newly available data***

Result 5 – Flexibility and autonomy must include dealing with uncertainty and incompleteness

Socio-technical systems are basically characterized as being incomplete - this is a reason for continuous transformation and evolution

→ Being incomplete is also relevant for AI-Solutions

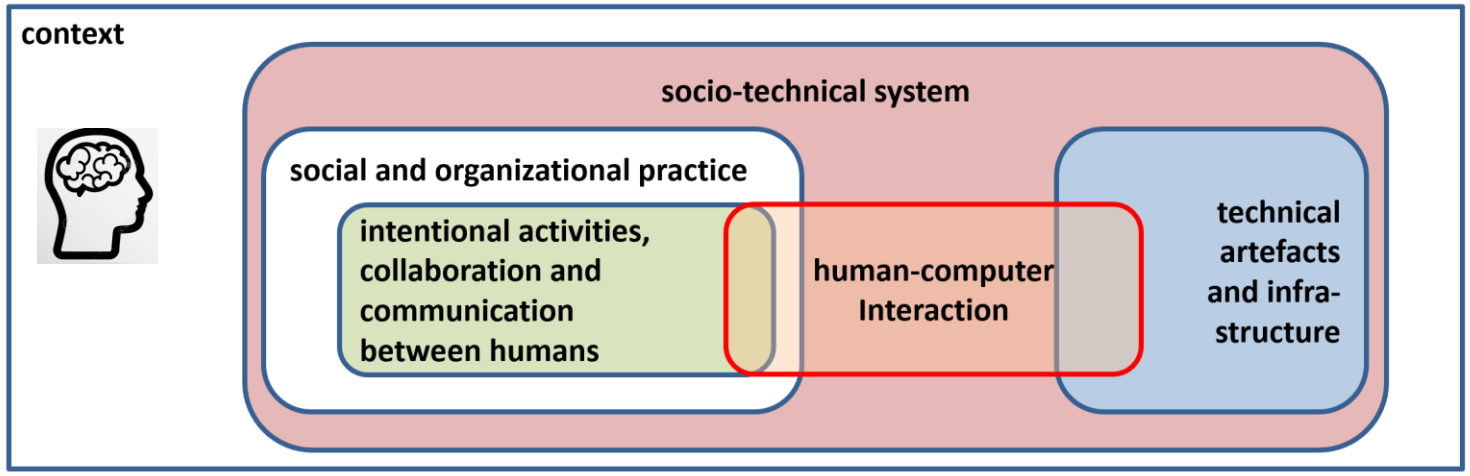
Accepting flexibility, incompleteness and constant evolution conflicts with safety and robustness

Result 6 – Fairness is not sufficiently addressed in the socio-technical discourse

Within the conventional STS Design,
Fairness is only realized **within** a socio-technical system
if it is reflected in the values of the actors of the system and thus, becomes part of their responsibility – but actors outside in its context are not considered.

→ Socio-technical systems have also to consider the interests of people outside the system who are affected by AI-based decision making

**Consequently, an additional socio-technical principle or heuristic such as 'value implantation' could be relevant
to implement fairness into social and organizational practices.**



Result 7 – Benefit must be balanced with effort

Users may be willing to sacrifice ease of use or efficiency if they experience an advantage in terms of other values instead.

→ people may be willing to exercise control and oversight when using AI – even if this causes inefficiencies, because being in control is a value in itself.

Trust calibration: The relation between feeling a need for control and trust building is not sufficiently discussed within STS-research.

Conclusion

The field of HCAI, responsible AI and ethical guidelines for AI on the one hand, and socio-technical design principles and heuristics on the other do not completely overlap and can benefit from each other.

Constant evolution, incompleteness → HCAI

Fairness, trust calibration → STS

Principles and guidelines cannot only regulate AI usage

But AI can also help to meet these principles and guidelines

Thank you for your attention!

thomas.herrmann@rub.de