Organizational issues in human-AI teaming: Governance and uncertainty management

Gudela Grote
Professor of Work and Organizational Psychology
ETH Zürich

I would not want to develop AI-based medical tools because I have too little control over the AI models.

(Data scientist in a software development company)

It never occurred to me that XAI could be (mis)used to hold AI users accountable.

(HCI researcher)

Overview of talk

- Control and accountability as fundamental concepts of governance

- Control and accountability in socio-technical design

- New challenges for aligning control and accountability for AI systems

- How to address the new challenges

- Avenues for research

ETH *zürich*

Control and accountability as fundamental concepts of governance

Control is defined as an actor's ability to achieve desired and avoid undesired outcomes.

Control entails influence over a current situation and sufficient understanding of the situation arising from transparency of ongoing processes and predictability of future states and outcomes.

Accountability: "relationship between an actor and a forum, in which the actor has an obligation to explain and justify his or her conduct, the forum can pose questions and pass judgement, and the actor may face consequences" (Bovens, 2007).

In corporate governance, "effectiveness involves the accountability of organizational decision-makers and the legitimacy of decisions about their economic and non-economic goals and values" (Aguilera et al., 2008).

Managerial/organizational control: Being able to align individuals' and groups' behavior with the goals of the organization (through making them accountable).

(Brehmer, 1992; Bridoux & Stoelhorst, 2022; Carver & Scheier, 1990; Frink et al., 2008; Green & Walsh, 1988; Hall et al., 2017; Hollnagel & Woods, 2006; Merchant & Otley, 2006; Ouchi, 1979; Sitkin et al., 2020; Skinner, 1996)

Implications for aligning control and accountability

o   Control enables actors to achieve desired and avoid undesired outcomes they are held accountable for.

o   Accountability motivates actors to act in alignment with the goals of a superordinate social system and is thereby a mechanism for managerial/organizational control.

o   Misalignment results from control without accountability or accountability without control – these two forms of misalignment are often connected when actors with control transfer accountability to actors without control.

o   Accountability and control may be aligned across actors or even across organizations.

o   Sharing of accountability and control is possible but requires handling of freerider and exploitation problems.

Control and accountability in socio-technical design

Which car would you prefer?

Effects of increasing automation

- Mix of qualitative overload and quantitative underload for human operators

- Human operators as stop gap for not yet automated functions

- Loss of human knowledge

- Misfit between accountability and control

Core question:
Automation or augmentation to keep human in the loop?

(Billings, 1997; Endsley, 2016; Hollnagel & Woods, 2006; Parker & Grote, 2022)

ETH *zürich*

# Ironies of automation: Crucial but impossible tasks for humans?

"The correct functioning of the train control system and the automatic traffic control system is to be monitored by the signaller. If necessary, he/she has to intervene manually.

During normal operation, no monitoring is necessary as long as the operational requirements are met.

In the case of disturbances or incidents, the notification of the required services and the required alarm procedures must be guaranteed."

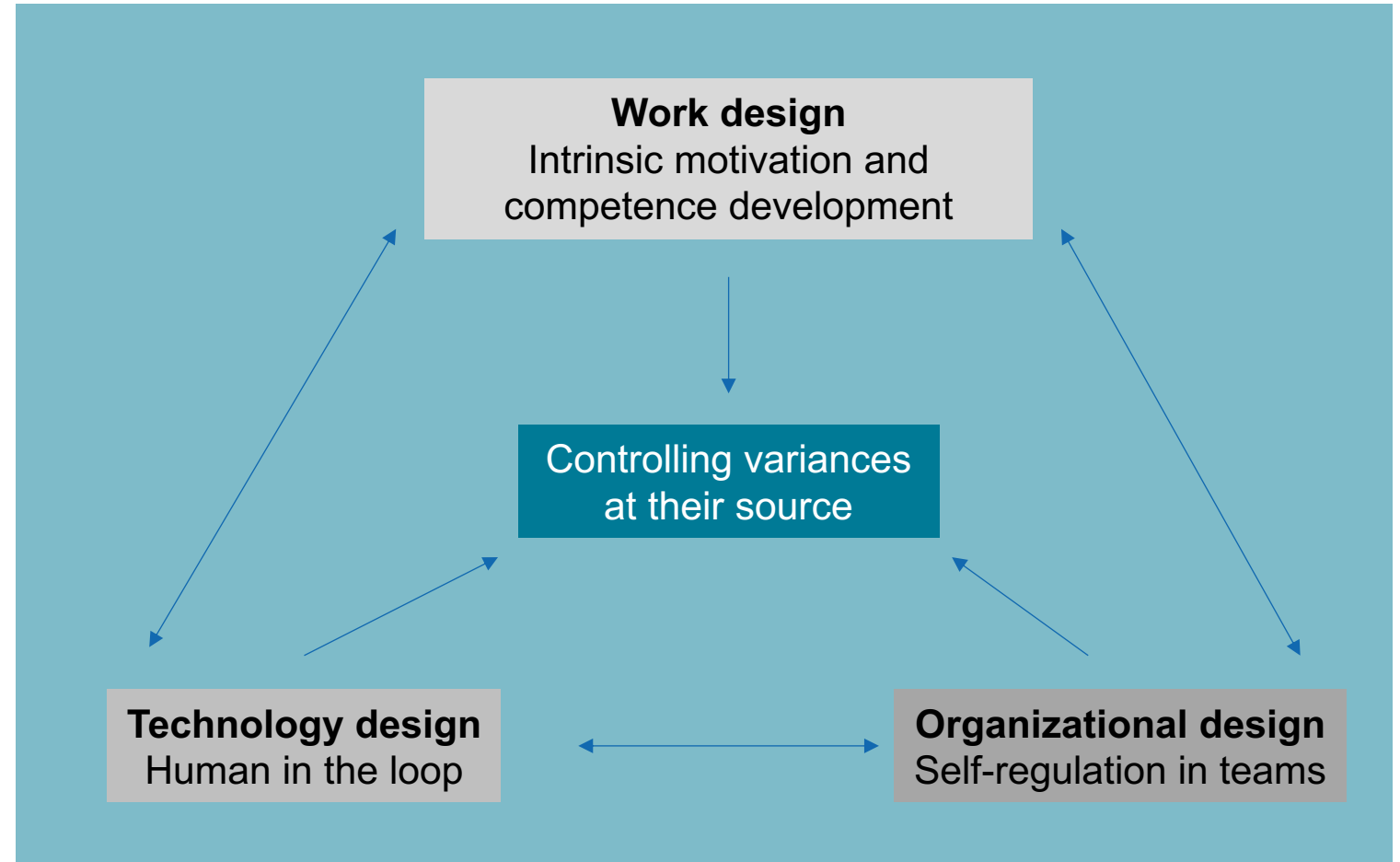(Excerpt from the rule book of a European railway company)

(Bainbridge, 1983)
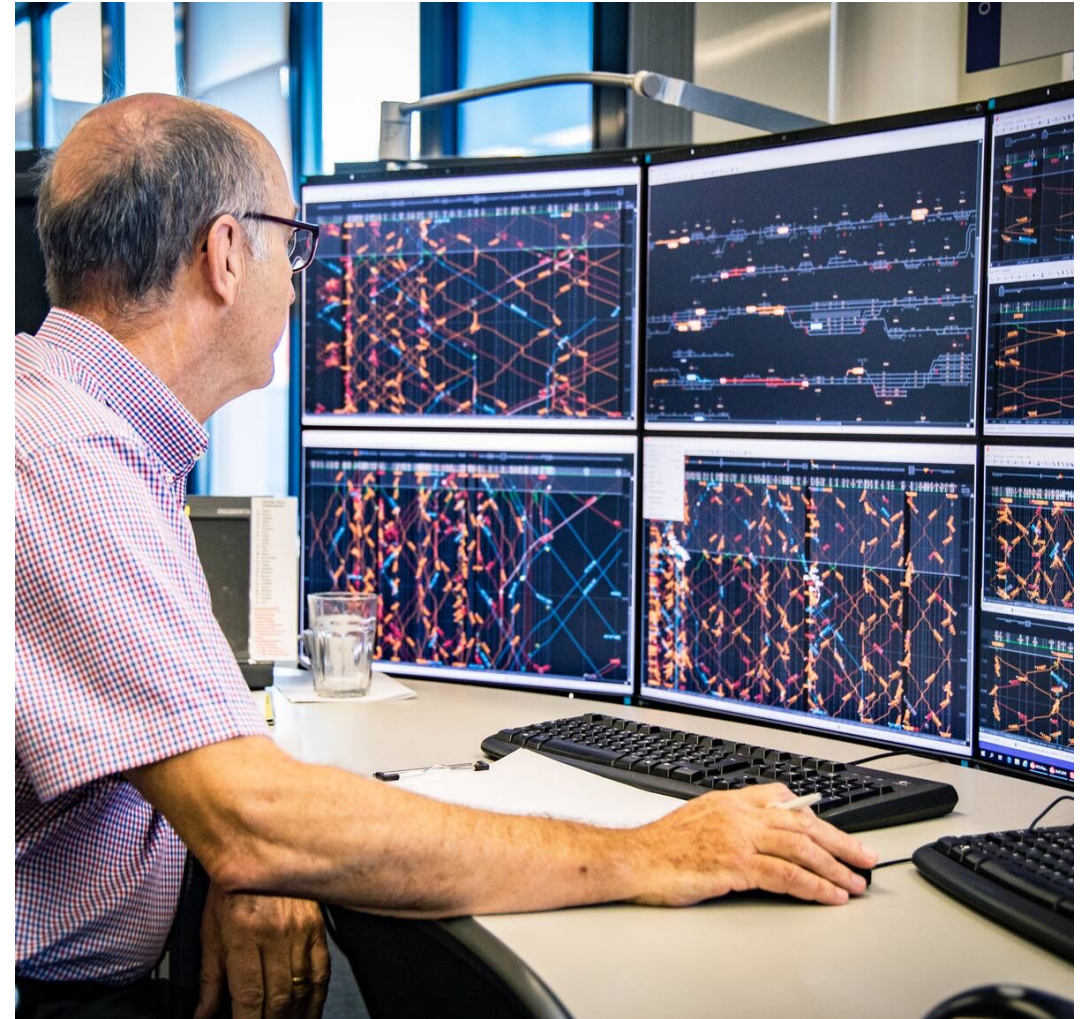
Complementarity between human and technology: Building on the strengths of both with human-in-the-loop design
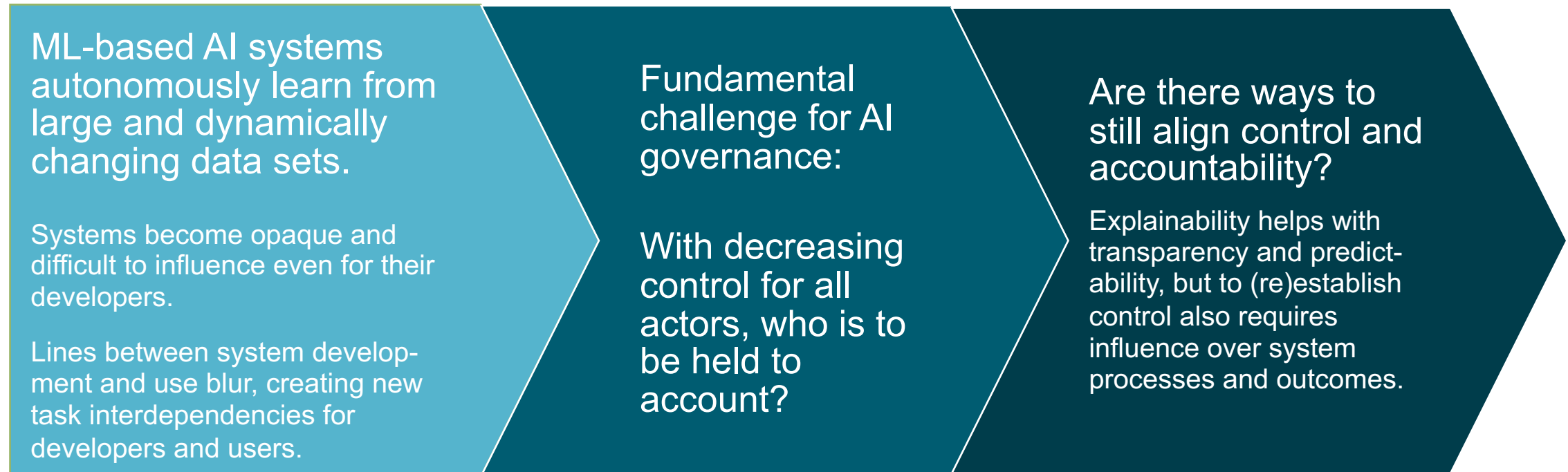
Supporting complementary design with the method KOMPASS:

- create a more holistic and shared design mindset among technology developers;

- foster systematic consideration of work design principles already in early phases of technology development;

- facilitate processes of continuous technology-work co-constitution.

Work design
Intrinsic motivation and competence development

Controlling variances at their source

Technology design
Human in the loop

Organizational design
Self-regulation in teams

(Billings, 1997; Grote et al., 1999, 2000; Hollnagel & Woods, 2006; Wäfler et al., 1999, 2003; Boos et al., 2013)

New challenges for aligning control and accountability for AI systems

ML-based AI systems autonomously learn from large and dynamically changing data sets.

Systems become opaque and difficult to influence even for their developers.

Lines between system development and use blur, creating new task interdependencies for developers and users.

Fundamental challenge for AI governance:

With decreasing control for all actors, who is to be held to account?

Are there ways to still align control and accountability?

Explainability helps with transparency and predict-ability, but to (re)establish control also requires influence over system processes and outcomes.

(Anthony et al., 2023; Asatiani et al., 2021; Berente et al., 2021; Castelvecchi, 2016; Faraj et al., 2018; Jacobides et al., 2021; Rudin, 2019; Slota et al., 2023; Wieringa, 2020)

Considerations for control-accountability alignment in AI development and use

(Grote, Parker, & Crowston, in prep.)

AI capabilities for autonomous adaptability	Low (e.g., programmed system; trained system with few parameters)	Medium (e.g., trained system with many parameters, but frozen)	High (e.g., continuously learning deep neural networks)
Control of AI outcomes	AI users are in control if the system is explainable and leaves final decision-making to them.	AI users are at best partially in control if the system is explainable and still leaves certain decisions to them. AI developers are in control by defining and maintaining an operating envelope for system use.	AI users have no control and fully rely on the system as part of their work tasks, possibly aided by some explanantions given by the system. AI developers are partially in control by intense testing and monitoring of system outcomes within an operating envelope.
Accountability for AI outcomes	AI users are accountable if conditions for their control have been established.	AI developers are accountable if conditions for their control have been established.	AI developers are accountable if conditions for their (partial) control have been established.
Control over AI system functioning	AI developers are in control if they have decision power over ML techniques in line with the chosen system functionality.	AI developers are in control if they have decision power over ML techniques in line with the chosen system functionality.	AI developers are partially in control by intense testing and monitoring of system functioning within an operating envelope.
Accountability for AI system functioning	AI developers are accountable if conditions for their control have been established.	AI developers are acountable if conditions for their control have been established.	AI developers are acountable if conditions for their control can be established; otherwise, senior management of developers is accountable.
Accountability for supporting organizational mechanisms	Senior management of users and developers are accountable for strengthening agency of AI users and developers.	Senior management of users are accountable for preventing control abuse by AI users. Senior management of developers are accountable for strengthening agency of AI developers.	Senior management of users and developers are accountable for endorsing organizational monitoring and feedback systems for continuous learning.

ETHzürich

Increasing uncertainty for both AI users and developers

A process model for control-accountability alignment in stakeholder negotiations (Grote, Parker, & Crowston, in prep.)

Involvement of stakeholders from full AI life cycle with shared governance

Integrative negotiations facilitated by extensive information sharing, perspective taking and accountability as shared norm

Control-accountability alignment for AI developers and users

Degree of autonomy and adaptivity of AI system

Possibility for control for users or developers

Appropriate assignment of accountability

Ability and motivation through aligned control and accountability

Risk mitigation by achieving desired and avoiding undesired outcomes

(Brett & Thompson, 2016; Bridoux & Stoelhorst, 2022; Chhillar & Aguilera, 2022; Falco et al., 2021; Galinsky et al., 2008; Hardy et al., 2020; Hwang et al., 2019; Jasperson et al., 2002; Lebovitz et al., 2021, 2022; Macrae, 2021; Thompson et al., 2010)

ETH zürich

One important avenue for future research

- Better understand how technology developers manage uncertainty in their daily work and how they can be best supported
    - tools (e.g., for risk assessment)
    - work design
    - senior management action

(Hagtvedt, 2019; Jain, 2023; Myers, 2023; Suchman, 2002)

ETH *zürich*

Thank you!

Contact: ggrote@ethz.ch

ETH zürich