Combining Human and Artificial Intelligence

Swiss Safety Committee (SSC), Meeting 2024-1

Olten, 5.9.2024

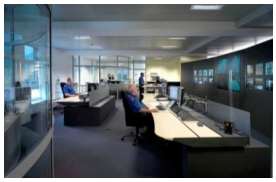
Prof. Dr. Toni Wäfler

University of Applied Sciences and Arts Northwestern Switzerland
School of Applied Psychology

Institute Humans in Complex Systems (MikS)

– Analysis, evaluation and design of complex systems
  – Increased technical and organizational complexity
  – Individuals and groups confronted with such complexity

– Objectives
  – Increase the reliability and safety of sociotechnical systems
  – Healthy humans and organizations



Outline

- Intelligence
- Amazing abilities of machines
- Humans are different
- Amazing abilities of humans
- Outlook: Combining man and machine
  HORIZON-Project AI4REALNET

Intelligence

Intelligence:
Ability to learn from experience, solve problems and use knowledge to adapt to new situations.
(Myers, 2005, S. 460)

Artificial intelligence:
Replication of human intelligence within computer science.
(Wikipedia, 2023)

Bild von Susanne Stöckli auf Pixabay

Focus: Experts

Bild von skyguide.ch

Outline

–Intelligence

–Amazing abilities of machines

–Humans are different

–Amazing abilities of humans

–Outlook: Combining man and machine
HORIZON-Project AI4REALNET

Enormous Increase in Hardware Performance

Performance doubles approximately every two years (cf. Moore's law)

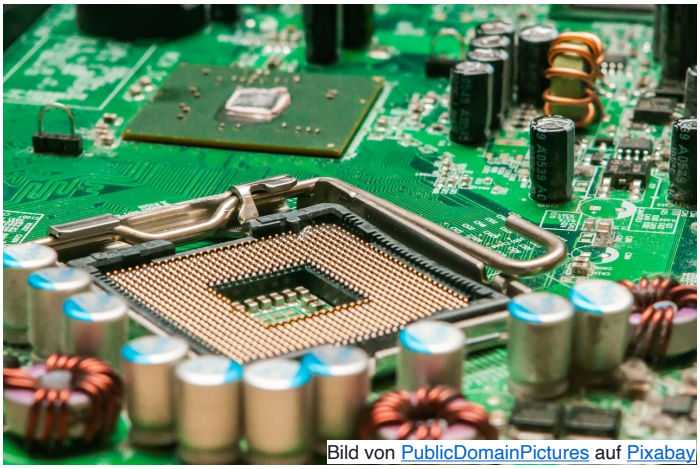


Bild von PublicDomainPictures auf Pixabay

Amazing Abilities of Machines

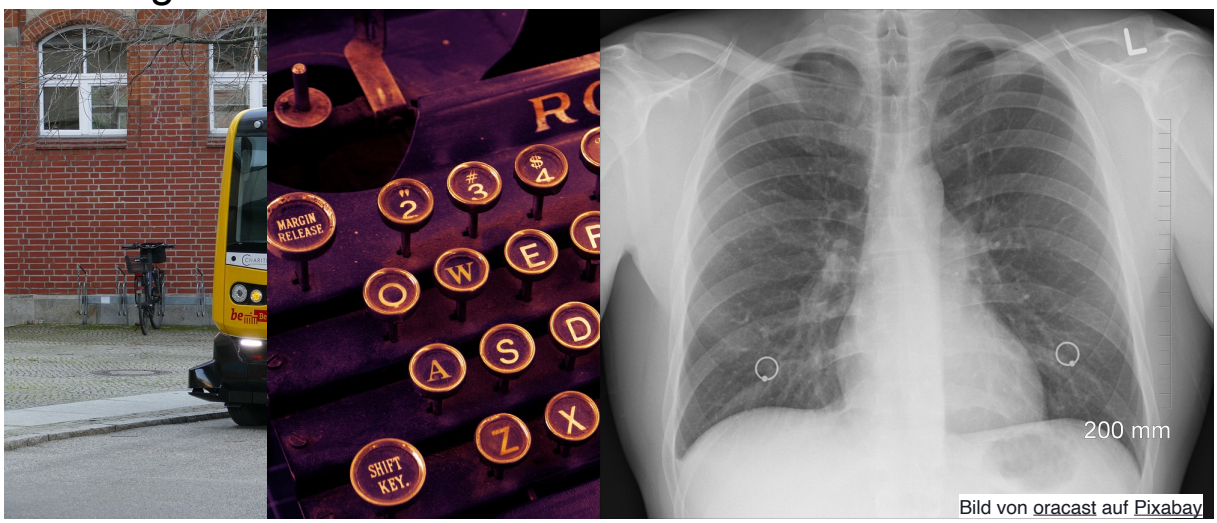


Bild von oracast auf Pixabay

Outline

–Intelligence
–Amazing abilities of machines
–Humans are different
–Amazing abilities of humans
–Outlook: Combining man and machine
HORIZON-Project AI4REALNET

Humans are Different

Object recognition vs.
patterns in pixels

Bild von forthdown auf Pixabay

Humans are Different

Object recognition vs. patterns in pixels

Understanding vs. data patterns

Bild von Sasin Tipchai auf Pixabay

University of Applied Sciences and Arts Northwestern Switzerland
School of Applied Psychology

Humans are Different

Object recognition vs. patterns in pixels

Understanding vs. data patterns

Value system (understanding of the world) vs. pursuit of goals?

Bild von Engin Akyurt auf Pixabay

Humans are Different

- Object recognition vs. patterns in pixels
- Understanding vs. data patterns
- Value system (understanding of the world) vs. pursuit of goals?
- Empathy (Theory of Mind) vs. inability to put oneself in the other person's shoes

Bild von PublicDomainPictures auf Pixabay

University of Applied Sciences and Arts Northwestern Switzerland
School of Applied Psychology

Humans are Different

- Object recognition vs. patterns in pixels
- Understanding vs. data patterns
- Value system (understanding of the world) vs. pursuit of goals?
- Empathy (Theory of Mind) vs. inability to put oneself in the other person's shoes
- Taking responsibility vs. functioning

Bild von Renno_new auf Pixabay

Outline

–Intelligence
–Amazing abilities of machines
–Humans are different
–Amazing abilities of humans
–Outlook: Combining man and machine
HORIZON-Project AI4REALNET

Expertise I: Humans' Experience = Tacit Knowledge

Experts recognise patterns:
They can „read" situations

Often an unconscious process

Mostly tacit knowledge: not explainable

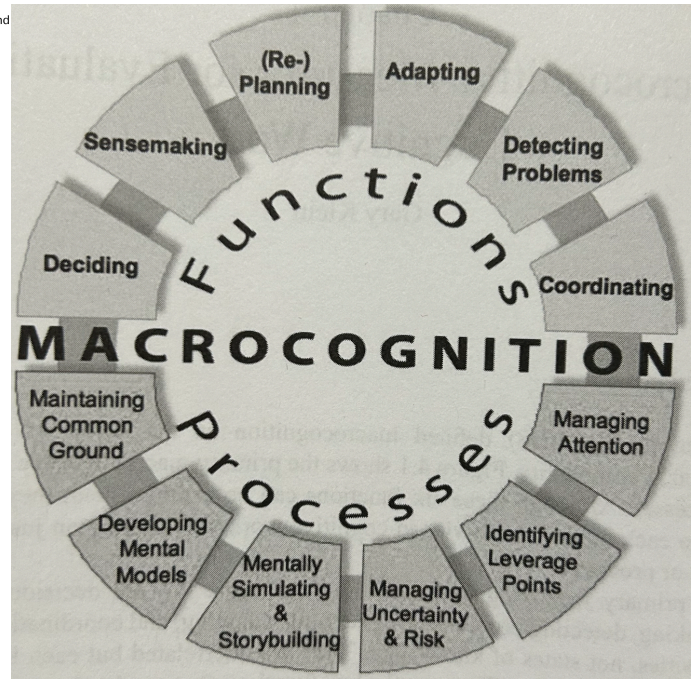
Prerequisite: Experience

- What is important about the situation?
- What needs to be taken into account?
- What needs to be done?

(Klein, 1993)

Bild von Bernd auf Pixabay

Macrocognition



(Klein, 2010)

Expertise II: Humans' Experience = Trained Skills

Many skills are also experience-based

Cannot be learnt theoretically

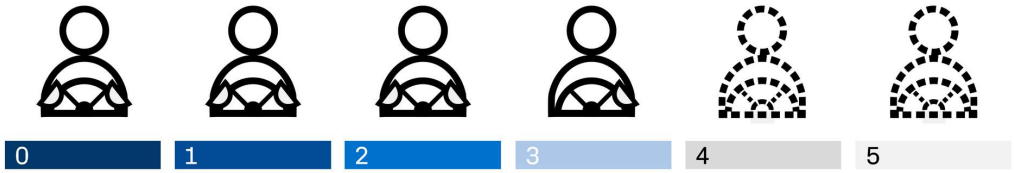
Need to be trained

Bild von Simon auf Pixabay

Outline

–Intelligence
–Amazing abilities of machines
–Humans are different
–Amazing abilities of humans
–Outlook: Combining man and machine
HORIZON-Project AI4REALNET

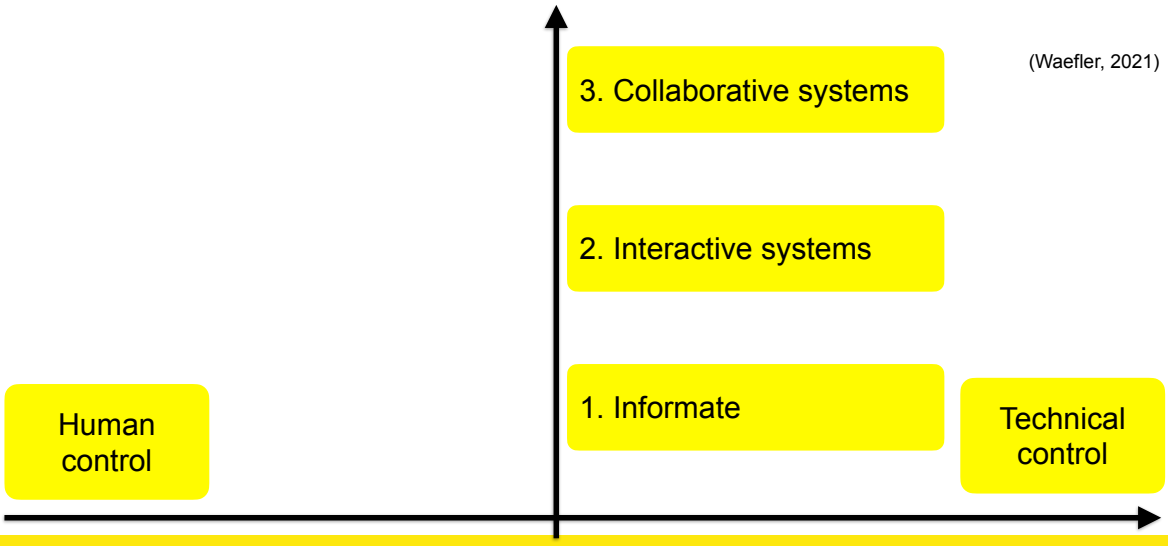
Classic View: Automation = Increasing Technical Autonomy



0	1	2	3	4	5
No automation	Driver assistance	Partial automation	Conditional automation	High automation	Full automation
Zero autonomy: the driver performs all driving tasks.	Vehicle is controlled by the driver, but some driving assist features may be included in the vehicle design.	Vehicle has combined automated functions, like acceleration and steering, but the driver must remain engaged with the driving task and monitor the environment at all times.	Driver is a necessity, but is not required to monitor the environment. The driver must be ready to take control of the vehicle at all times with notice.	The vehicle is capable of performing all driving functions under certain conditions. The driver may have the option to control the vehicle.	The vehicle is capable of performing all driving functions under all conditions. The driver may have the option to control the vehicle.

(credit-suisse.com)

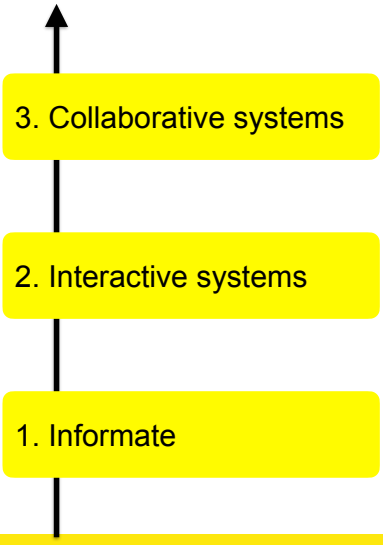
Alternative View: Increasing Cooperation of Humans and Technology



(Waefler, 2021)

(Wäfler, 2020)

Human Roles



Verification of AI: Verify AI-based decision proposals

Improving AI: Identifying weaknesses and biases in AI-based decisions

Learning from AI: Learning from the AI, which also expands the horizon of human experience

Taking responsibility and compliance with regulatory and ethical standards

(Samek et al., 2017)

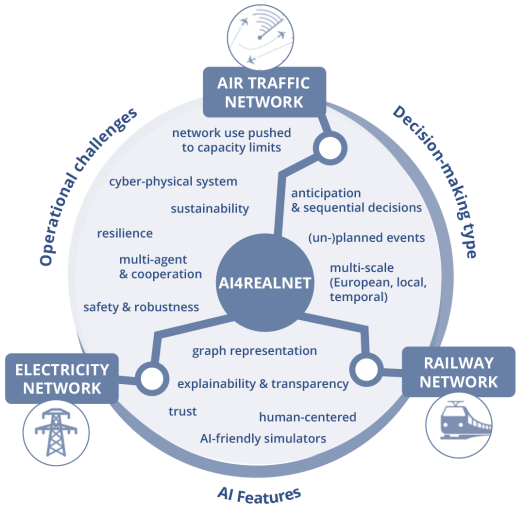
"To trust it, they need to understand what the machines are doing." (Davide Castelvecchi)

AI4REALNET.EU

AI4REALNET has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101119527

ai4realnet.eu

Focus on critical infrastructures



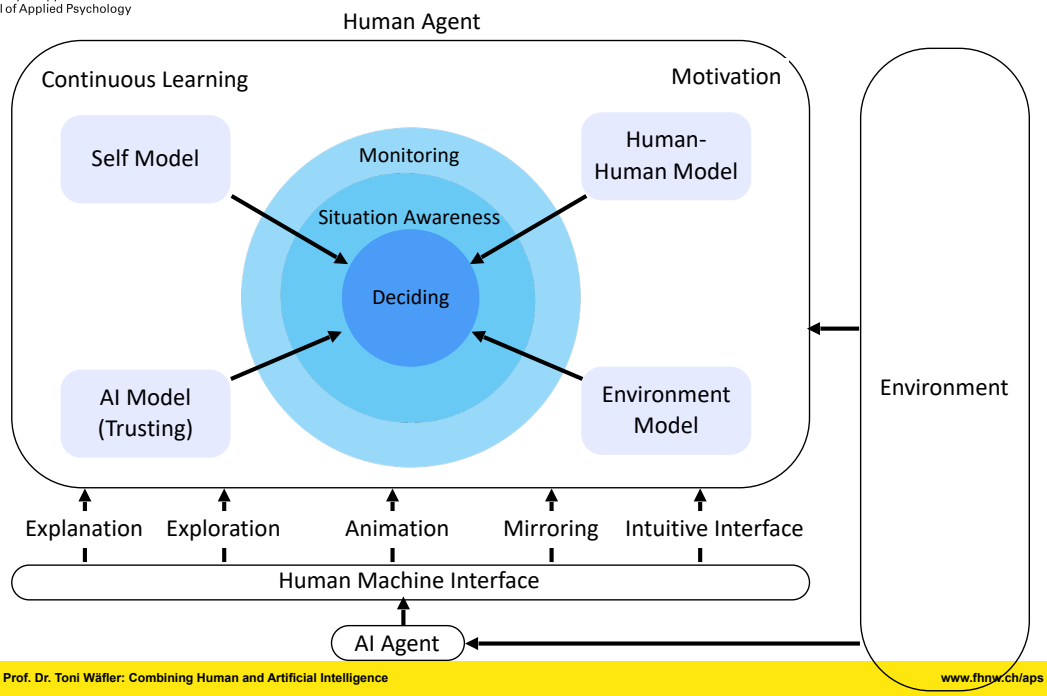
Our core elements are:

1 **AI algorithms mainly composed by supervised and reinforcement learning.**
Unifying the benefits of existing heuristics, physical modelling of these complex systems and learning methods, as well as a set of complementary techniques to enhance transparency, safety, explainability and human acceptance.

2 **Human-in-the-loop decision making for co-learning between AI and humans.**
Considering integration of model uncertainty, human cognitive load and trust.

3 **Autonomous AI systems relying on human supervision.**
Embedded with human domain knowledge and Safety rules.

AI4 REALNET | ai4realnet.eu ai4realnet.eu



Summary

- Humans and artificial intelligence are very different
- Artificial intelligence can (partially) emulate human intelligence
- Based on their different strengths and weaknesses, humans and artificial intelligence can complement each other

References

- Mitchell, M. (2019). Artificial Intelligence. New York: Farrar, Strauss and Giroux.
- Klein, G. A. (1993). A recognition-primed decision (RPD) model of rapid decision making. Decision making in action: Models and methods, 5(4), 138-147.
- Russell, S. (2019). Human Comaptible. Artificial Intelligence and the Problem of Control. Penguin Books.
- Samek, W., Wiegand, T. & Müller K.R. (2017). Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models. ITU Journal: ICT Discoveries, Special Issue The Impact of AI on Communication Networks and Services, 1, 110.
- Waefler T. (2021). Progressive Intensity of Human-Technology Teaming. Proceedings of the 5th International Virtual Conference on Human Interaction and Emerging Technologies, IHIET 2021, August 27–29, 2021, France, pp. 28-36.
- Wäfler, T. (2020). Gebildeter und vernetzter Mensch: Vier Thesen zur soziotechnischen Gestaltung der Zukunft. Journal Psychologie des Arbeitshandelns. 13(2), S. 5-21.

Thank you for your attention!

toni.waefler@fhnw.ch
toniwaefler.ch