VTT Technical Research Centre of Finland

Key Human Factors Concepts of Artificial Intelligence Awareness

Karvonen, Hannu; Heikkilä, Eetu; Wahlström, Mikael

Published: 12/12/2018

Document Version
Publisher's final version

Key Human Factors Concepts of Artificial Intelligence Awareness: Transparency, Communication, and Trust

Hannu Karvonen, Eetu Heikkilä, Mikael Wahlström
VTT Technical Research Centre of Finland Ltd

INTRODUCTION

- Increase of AI usage (e.g., machine learning, ML) → More awareness -related interaction challenges for the users of these systems.
- **Situation awareness (SA), automation awareness (AA), and artificial intelligence awareness (AIA)** as important concepts
  - Focus of this work: **user AI awareness regarding ML processes**
  - Related human-AI interaction issues:
    1. **System transparency**
    2. **Computer-human communication**
    3. **Appropriate trust**
- Key question: **How these phenomena relate to each other?**

KEY CONCEPTS AND ISSUES

- Definition of AIA: ***The human's perception of the current decision made by the AI, his/her comprehension of this decision, and his/her estimate of the decision(s) by AI in the future.***

Key human factors concepts concerning human-AI interaction:

A. Transparency:

- *How transparent is the functioning of the AI?*
  - i. **Black box of ML → transparent and clear ML process**
    - ➢ The probabilities of the predictions in conducted tasks could be presented to the user to enable the estimation of their reliability.
  - ii. **Design implications:**
    - ➢ Understandable explanations of the ML process
    - ➢ Reasons behind certain decisions and results
    - ➢ Simplifications and illustrative visualizations of used algorithms

B. Communication:

1. *The way in which the AI system communicates its functioning, intentions, capabilities, and limitations to the user*
2. *The possibilities of the AI to understand human communication.*
   - i. **E.g., simplifications and clear visualizations are needed**
     - ➢ Explicit information visualization techniques,
     - ➢ Multimodal output user interfaces
     - ➢ Interaction technologies that are natural for humans
   - ii. **The communication of humans understood by the computer: natural ways of humans to communicate** (e.g., voice/gestures)

C. Appropriate trust:

- *Can the user trust the AI, based, on the knowledge about, e.g.,*
  1. *The capabilities of the AI-based system,*
  2. *The quality and relevancy of the used data, or*
  3. *Has the system learned the required skills w/o becoming biased.*
  - i. **Calibration of trust into an appropriate level (see Fig 1)**
    - ➢ User has to have a clear idea of the capabilities of the used algorithms and what kind of data has been used in the ML
    - ➢ Mitigating the problematic effects of overtrust or distrust in AI.

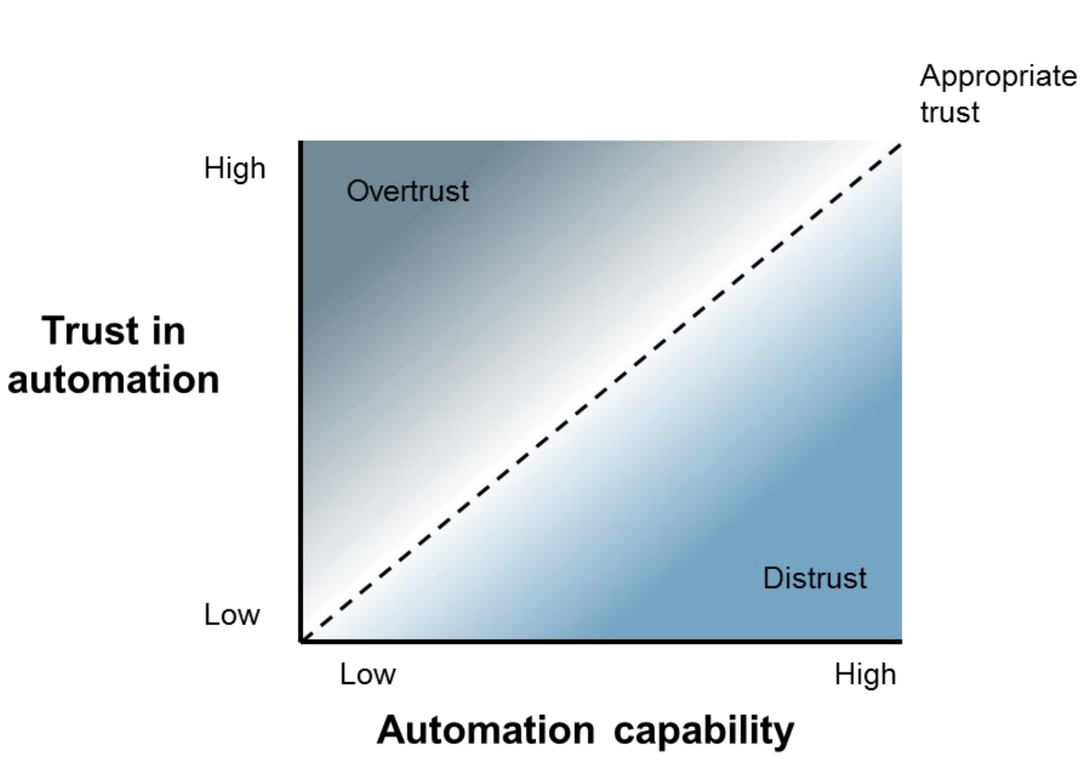


Figure 1. *Appropriate trust in automation. Modified based on Lee & See (2004).*

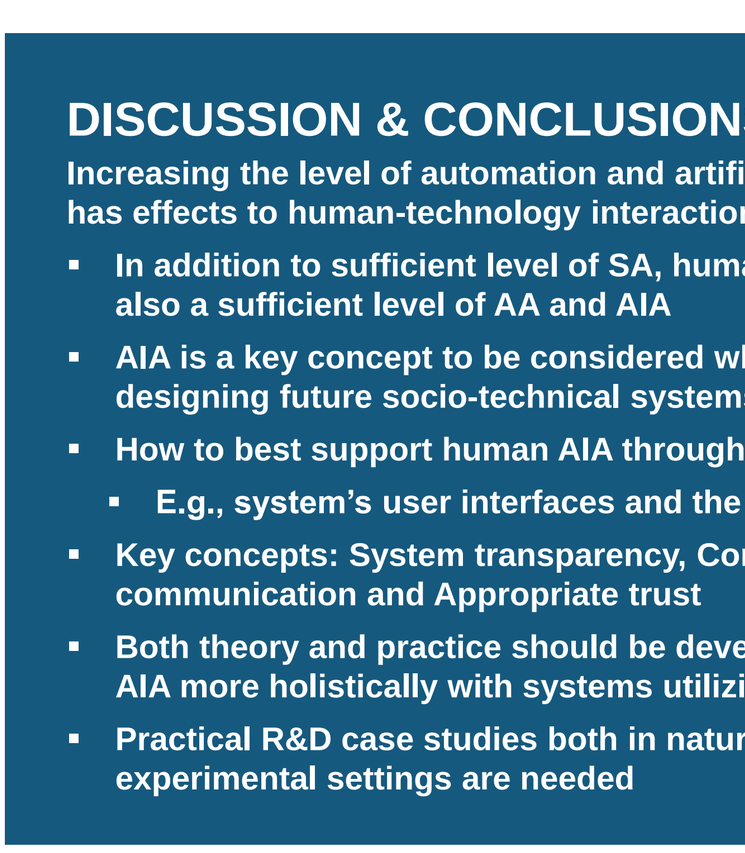
ARTIFICIAL INTELLIGENCE AWARENESS AND AFFECTING MATTERS

AIA is based on the ways for the human to gain understanding about the AI's functioning, which is affected by e.g.,

1. AI system's user interfaces
2. Provided AI-specific training and education
3. General knowledge of the principles of computer systems and AI
4. Momentary high level of cognitive workload of the human
5. Subjectively experienced level of complexity of the AI system.

TAXONOMY OF SA, AA, AND AIA

- We suggest that the discussed awareness-related phenomena enclose each other in the following way:
  - **AA encloses AIA while SA encloses these both**
- **To illustrate the taxonomy, we present Fig 2, in which the stages of awareness are based on the three-level SA model by Endsley.**
- The circular form of the illustration in Fig 2 refers to the formation of SA, AA and AIA as a continuous process instead of a linear one.
  - Neisser's action-perception cycles in human activity.
  - In line with the Situated Cognition perspective, emphasizing how current awareness of a situation affects the process of acquiring and interpreting new awareness in an on-going cycle.



Explanations for the used codes:

- ***1A**: Perception of the current situation*
- ***1B**: Comprehension of the current situation*
- ***1C**: Estimate of the future situation*
- ***2A**: Perception of the current status of the automation*
- ***2B**: Comprehension of the current status of the automation*
- ***2C**: Estimate of the future status of the automation*
- ***3A**: Perception of the current AI-based decision by the system*
- ***3B**: Comprehension of the current AI-based decision basis of the system*
- ***3C**: Estimate of the system's AI-based decision(s) in the future and their basis*

Figure 2. *The Awareness Circles of SA, AA and AIA. The Figure is meant to reflect the relationship between 1. Situation Awareness, 2. Automation Awareness, and 3. AI Awareness.*

DISCUSSION & CONCLUSIONS

Increasing the level of automation and artificial intelligence has effects to human-technology interaction:

- **In addition to sufficient level of SA, human need to achieve also a sufficient level of AA and AIA**
- **AIA is a key concept to be considered when studying and designing future socio-technical systems.**
- **How to best support human AIA through design choices?**
  - **E.g., system's user interfaces and the provided training**
- **Key concepts: System transparency, Computer-human communication and Appropriate trust**
- **Both theory and practice should be developed to consider AIA more holistically with systems utilizing ML**
- **Practical R&D case studies both in naturalistic and experimental settings are needed**

www.vttresearch.com

Contact
Hannu Karvonen
hannu.karvonen@vtt.fi