Should we (dis)trust robots? Developing responsible AI using cognitive and affective human-robot trust

Katharina Weitz1,* and Elisabeth André1

1Human-Centered Multimedia, Department of Computer Science, Augsburg University, Universitätsstraße 6a, Augsburg, Germany
*Corresponding author: katharina.weitz@informatik.uni-augsburg.de

In the last 10 years a moderate but continuous increase in the field of robotics can be seen [6], among which personal service robotics have the highest expected growth rate [2]. Personal service robots can enrich people's lives not only by providing physical support (e.g., as support for housework) but also by addressing psychological aspects (e.g., attention & caring). When using robots in private environments, a lack of trust has been observed. For example Reich-Stiebert and Eyssel [5] have shown that social acceptance for robots is often reserved. According to Lewis and Weigert [4] trust can be divided into cognitive and affective aspects. In the context of human-robot trust, cognitive trust can be seen as mental attributes, reasons and arguments of a person towards an agent, whereas affective trust describes the feeling of a person towards an agent [3].

In the ongoing doctoral thesis, affective and cognitive aspects of trust and distrust in human-robot interaction are investigated to develop a transparent, predictable and comprehensible system [7]. In a first step, different explainable artificial intelligence methods [1] are tested and analysed in the context of human-robot trust. The results will be used to create a responsible AI system that improves confidence in robots. Improvement of confidence should not be synonymous with a lack of questioning the actions and decisions of robots. Instead, it is aimed to enable humans to make decisions that are not dominated by fear.

References