Authors:
Cristina Zaga, Human-Centred Design Group, University of Twente
Nazli Cila, Digital Life Centre group, Amsterdam University of Applied Science
Maria Luce Lupetti, AITech group, Delft University of Technology

Title: *Towards responsible design of AI things: the role of agency as a relational quality in human-agent interactions*

Abstract: Digital technologies impact human behavior through "mere" if-this-then-that interaction schemas, while AI-Things (i.e., products and services that are imbued with intelligence and autonomy via AI) engage in relationships that are traditionally considered as exclusive to humans, unveiling uncharted forms of human-technology relations. The way AI-Things perceive, reason, and act to communicate with humans not only has an effect on the natural human tendency of attributing agency to things that behave, but it enriches human-agent interactions (HAI) of relational qualities. As humans and AI-Things become ontologically inseparable, notable socio-technical and moral implications arise. For instance, a companion robot can turn into a deceptive agent. The robot leveraging its human-like behaviors tries to convince people that it is a friend. In turn, people may delegate too much control to the robot that deceptively behaves like a human, disclosing personal information, or delegating their decisions to the robot.

How can we account for the socio-cultural, and moral, implications of our relations with AI Things? Ethics of Technology and STS studies provide various theoretical frames to examine the societal controversies around AI applications, but theories are often difficult to translate into actionable strategies to develop AI products. In this regard, we argue that design research is uniquely positioned to advance the state-of-the-art in HAI by bridging the gap between socio-ethical and technical disciplines. Our talk illustrates how design research offers the methodological frame to include post-human values and bring a relational conception of agency at the center of the responsible design of AI-Things.

Keywords: relationality; agency; human-agent interaction; responsible design.