Abstract

A Participatory Value-Sensitive Model of AI Assistant as an Opportunity for Systematic Risk Mitigation

There are a number of risks associated with the use of artificial intelligence systems that act as assistants, which directly interfere with human autonomy and affect self-determination. These technologies significantly affect life experiences and, consequently, a person's ability to fulfil his or her desires and goals. Although the companies that develop these products and the legislators who regulate them take into account the social, legal and ethical risks, in setting standards that mitigate them they are less focused on the role and potential contribution of the users themselves in the process of mitigating the risks that directly affect them. This dissertation therefore advances the current state of knowledge by providing a proposal for a model of the interface between AI assistants and their technological ecosystem in the form of a participatory model that would directly invite the user to actively participate in the modification of virtual models in the form of digital twins, thereby opening up new possibilities for significantly mitigating the risks associated with their use. This approach to reconciling security measures on the part of developers, legislators and users is grounded in the literature on ethics. Specifically, by studying the relationships between design, human participation and values, this thesis highlights how human and societal values are reflected in active user participation, which provides opportunities for better alignment between human values and technology.

Key words: AI assistant, user, risks, value sensitive design, digital twin, participation, ecosystem