Anne Kaun[1] and Anu Masso[2]

[1]Södertörn University
[2]Tallinn University of Technology

Towards a Theory of Basic Values in Artificial Intelligence: Comparative Factor Analysis in Estonia, Germany, and Sweden

Bakgrund/Frågeställning

There is increasing attention for ethical issues and values in artificial intelligence (AI) design and deployment. However, we do not know how those values are embedded in the artificial artefacts and how they are perceived by the population groups. Based on the prior theoretical literature on ethical principles and (moral) values in AI, we designed an original survey instrument, including 15 value components and an original scale to estimate the individuals' importance of these values. The article is based on the representative population survey conducted in Estonia, Germany, and Sweden (n=4501), representing diverse socio-cultural welfare systems with diverse automation experiences. The exploratory factor analysis has revealed four underlying elements of values by the participants concerning the design and use of artificial intelligence: techno-social security, socio-cultural adaptability, data justice and social welfare. The comparison has revealed some embedded values being more universally valued among specific socio-economic groups across the three countries like data justice, others being more inherent to country context (residents in Sweden and Estonia value more techno-social security and social welfare, whereas in Germany the deficiency regarding all the underlying value dimensions was expressed) or social groups (women and older age groups value techno-social security generally more). The analysis of the associations reveal that higher valuation of techno-social security and social welfare is related to higher trust in public institutions in general, or data owned, collected, and used by public institutions. At the same time, techno-social security associated with lower agreement with automation in public sector and lower agreement with data justice principles; to express it differently, the higher valuation of techno-social security leads to lower agreement with data justice principles. Based on the exploratory factor analysis and inspired by the theory of human values, a framework to conceptualize the basic values in artificial intelligence is suggested.

Metod och Resultat

Konklusion