Understanding Risks: The Impact of Risk Perception on Trusts and Acceptability of AI-Systems.

Background

As stated in multiple guidelines and upcoming european regulations, transparency and explainability are a requirement for ethical by design AI-powered technologies. Explainability of AI can be understood from various angles: interpretability for the developers, and understandability for the users. It is this second side that interests us in this research, as a step towards an accountability framework for AI-systems.

Expected goals

Drawing upon past literature, and survey designed data collection, this thesis will aim at identifying users' risk perception based on ethics principles, and the law. Additionally, acceptability and trusts towards AI depending on different risk types and levels will need to be assessed. Finally, a survey-based evaluation of users' acceptance and trusts in AIs depending on different AI caracteristics and risks will be lead.

Research Questions

- Which risks in AI systems are perceived as acceptable by the user?
- Which categories and levels of risks impact users' trust and acceptance of AI-Systems most?

Recommended literature

Abdul, A., Vermeulen, J., Wang, D., Lim, B. Y., & Kankanhalli, M. (2018, April). Trends and trajectories for explainable, accountable and intelligible systems: An hci research agenda.
In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1-18).

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689-707.

Xu, W. (2019). Toward human-centered AI: a perspective from human-computer interaction. *Interactions*, *26*(4), 42-46.

Details

Supervisor: Auxane Boch
Starting date: as soon as possible

Contact

If you are interested, please contact Auxane Boch (**auxane.boch@tum.de**)

We are looking forward to your application!