Values for Responsible AI-Based Decision-Aid for Fire Services

Tina Mioch[1,2] [0000−0003−2457−5872], Huib Aldewereld[1] [0000−0001−6339−3241], Luciano Cavalcante Siebert[2] [0000−0002−7531−3154], Stefan Leijnen[1] [0000−0002−4411−649X], and Mark A. Neerincx[2] [0000−0002−8161−5722]

[1] University of Applied Sciences Utrecht, The Netherlands
[2] Delft University of Technology, The Netherlands
{tina.mioch, huib.aldewereld, stefan.leijnen}@hu.nl,
{l.cavalcantesiebert, m.a.neerincx}@tudelft.nl

Abstract. Artificial Intelligence (AI) systems are more and more being introduced into first response, as they can lead to more effective incident response; this introduction needs to be done responsibly to avoid negative consequences. Generic guidelines and legislative measures for responsible AI-development and deployment have been proposed, but more in-depth analysis is required to understand the nature of responsible AI within this domain. For responsible development, AI systems should support relevant human values; however, it is not clear which values these are. We empirically investigate relevant values that are affected by the AI-based Decision Aid (AIDA), a decision support system under development for fire services in the Netherlands. For this interdisciplinary research, we held 10 expert group sessions in which we discussed expected impact of AIDA to provide the necessary foundation for responsible AI development in the field. This paper presents preliminary insights and steps forward.

Keywords: Values · Fire Services · Responsible AI · Decision-Support.

1 Introduction

AI systems have been introduced into first response (FR), e.g., for fire detection [8] and real-time surveillance of the incident scene [6]. This introduction of AI systems should be done responsibly to avoid negative consequences such as overreliance on systems. Several guidelines have been developed for the responsible development of AI systems [2, 3] as well as EU legislation [1]. An important principle of these initiatives is that the context in which AI systems are applied largely determines the ethical, legal, and societal impact and how to deal with this impact responsibly. In our research project, we develop a generic framework of ethical aspects of AI systems for FR to be used as a toolbox towards responsible AI design, development, and deployment. Our research is inspired by the Value Sensitive Design methodology, which accounts for human values throughout the design process [4].

In a previous study [7], we identified stakeholders and a first set of values regarding AI systems for FR. However, the applicability of the values depends on the specific AI system and context of use [9]. In this research, we further develop and deepen the

Table 1. Examples of impact identified for the video analysis module with its impacted values.

Examples of identified impact	**Impacted value**
Interpreting information becomes more difficult; Dependence on technology.	Autonomy
Dispatcher role might become less relevant.	Identity
Better situation awareness; More PTSD because of availability of incident videos.	Well-being
Negative consequences for personnel when AI information is not acted on.	Accountability

identified relevant values for a concrete use case, the AI-based Decision Aid (AIDA). AIDA is a decision-support system for fire services that structures and analyzes data, e.g., drone camera pictures and historic incident data. We focus on two modules: a *video analysis module* that analyzes drone video streams and detects humans [5] and a *material prediction module* that, based on correlations found in past incident data, predicts the needed fire services material for an incident. As this is work in progress, we describe the method, preliminary insights and next steps.

2 Method, preliminary results and next steps

To identify relevant values of different stakeholders for AIDA, we conducted 10 expert group sessions with in total 40 employees of the fire services, of which 17 (head) incident commanders, 19 dispatchers, 3 fire fighters, and 1 emergency coordinator. The participants identified positive and negative impact of the two AIDA modules. The identified impact of AI systems was analyzed through thematic analysis and mapped onto affected underlying values. We coded inductively as a way to enter the data analysis with a more complete, unbiased look at the themes throughout the data.

In Table 1, examples of impact and corresponding value identified for the video analysis module are presented. Note that this is a selection of impact participants mentioned, as we are still in progress of analysing the results. The following aspects can be noted on the basis of the preliminary results: (1) several of the previously identified values are implicated by either of the two AI modules, e.g., *Identity* and *Well-being*; (2) impact on values such as *Autonomy*, *Trust in AI*, and *Accountability* has been mentioned for both modules. As next steps, we will further analyse the results and integrate them into the ethical framework, together with values found from normative sources and results of a scoping literature review on ethical aspects of AI systems applied to FR. The framework will give insight into the relevant ethical aspects to take into account when developing AI systems for FR. By achieving a more contextualized understanding of the relevant values, we can design systems that account for these values, leading to more effective systems with less harmful impacts.

Acknowledgments. We thank the first responders of *GB*, *VRR*, and *VRZHZ* for their participation. This research was supported by the HU through a '*promotievoucher*'.

References

1. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance) (Jun 2024), http://data.europa.eu/eli/reg/2024/1689/oj/eng, legislative Body: CONSIL, EP
2. AI HLEG: Ethics Guidelines for Trustworthy AI. European Commission (2019), https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419
3. ECP: Artificial Intelligence Impact Assessment (English version) (Jan 2019), https://ecp.nl/publicatie/artificial-intelligence-impact-assessment-english-version/, publication Title: ECP \textbar Platform voor de InformatieSamenleving
4. Friedman, B., Hendry, D.G.: Value sensitive design: Shaping technology with moral imagination. Mit Press (2019)
5. Heemskerk, J., Mioch, T., Maathuis, H., Aldewereld, H.: Long-range human detection in drone camera images. ISCRAM Proceedings **21** (2024)
6. Martinez-Alpiste, I., Golcarenarenji, G., Wang, Q., Alcaraz-Calero, J.M.: Search and rescue operation using uavs: A case study. Expert Systems with Applications **178**, 114937 (2021)
7. Mioch, T., Aldewereld, H., Neerincx, M.A.: Empirical investigation of values affected by ai systems for fire services. ISCRAM Proceedings **21** (2024)
8. Talaat, F.M., ZainEldin, H.: An improved fire detection approach based on yolo-v8 for smart cities. Neural Computing and Applications **35**(28), 20939–20954 (2023)
9. de Wet, J., Wetzelhütter, D., Bacher, J.: Revisiting the trans-situationality of values in Schwartz's Portrait Values Questionnaire. Quality & Quantity **53**(2), 685–711 (Mar 2019). https://doi.org/10.1007/s11135-018-0784-8, https://doi.org/10.1007/s11135-018-0784-8