Human Values for Responsible Decision-Support for Fire Services

Mioch, T.; Aldewereld, Huib; Neerincx, M.A.

Publication date
2024

Document Version
Final published version

Published in
Proceedings of the 21st ISCRAM Conference

Human Values for Responsible Decision-Support for Fire Services

Tina Mioch*
University of Applied Sciences Utrecht
Delft University of Technology
tina.mioch@hu.nl

Huib Aldewereld
University of Applied Sciences Utrecht
huib.aldewereld@hu.nl

Mark A. Neerincx
Delft University of Technology
M.A.Neerincx@tudelft.nl

ABSTRACT

Artificial Intelligence systems are more and more being introduced into first response; however, this introduction needs to be done responsibly. While generic claims on what this entails already exist, more details are required to understand the exact nature of responsible application of AI within the first response domain. The context in which AI systems are applied largely determines the ethical, legal, and societal impact and how to deal with this impact responsibly. For that reason, we empirically investigate relevant human values that are affected by the introduction of a specific AI-based Decision Aid (AIDA), a decision support system under development for Fire Services in the Netherlands. We held 10 expert group sessions and discussed the impact of AIDA on different stakeholders. This paper presents the design and implementation of the study and, as we are still in process of analyzing the sessions in detail, summarizes preliminary insights and steps forward.

Keywords

Values, Fire Services, Responsible AI, Decision-Support.

INTRODUCTION

In the last years, AI and data-driven systems are more and more being introduced into first response. Examples are real-time detection, monitoring, and analysis of threats and hazards (e.g., fire detection by means of machine learning algorithms (Talaat & ZainEldin, 2023)), real-time, continuous surveillance of the incident scene by means of drones or ground robots (e.g., by means of a machine-learning-based object detection system (Martinez-Alpiste et al., 2021)), and monitoring and analyzing physiological data of first responders (FRs) to detect (health) threats (e.g., by monitoring heart rate variability to determine current stress levels (Meina et al., 2020)). The development and deployment of AI systems raises new ethical issues (Stahl, 2021) and should be done responsibly, taking ethical, legal, and societal aspects into account. In the last years, many research programs have been started to further research how AI systems can be designed and built responsibly, both by industry (e.g., Google[1] and IBM[2]) and academics (e.g., Harvard[3]); also, law and regulations on responsible AI are planned (e.g., the EU Act (European Commission, 2021)). In addition, several guidelines have been developed for the responsible application of AI systems (AI HLEG, 2019; ECP, 2019). An important principle of all these initiatives is that the context in which AI systems are applied largely determines the ethical, legal, and societal impact of these applications and how to deal with this impact responsibly. To further investigate the responsible application of AI systems for first response, we have started a research project to set up a generic framework of ethical aspects of AI systems for first response. Our research is inspired by the Value Sensitive Design (VSD) methodology, which accounts for human values

*corresponding author
[1]https://research.google/research-areas/responsible-ai
[2]https://www.ibm.com/impact/ai-ethics
[3]https://cyber.harvard.edu/topics/ethics-and-governance-ai

throughout the design process (Friedman & Hendry, 2019). In a previous study (Mioch et al., 2024), first steps were made towards this generic framework by identifying relevant stakeholders and identifying and analysing a first set of human values of these stakeholders regarding AI systems for fire services. These general values have been identified to be important for this domain and should be taken into account when developing AI systems for FR. However, the applicability of the value set depends on the specific AI system and context of use and should be evaluated for each application.

In this research, we further develop and deepen the identified relevant human values for a concrete use case, the development of a specific AI application for the fire services, the AI-based Decision Aid (AIDA). AIDA is a decision-support system that is currently under development by the Dutch Safety Region *Rotterdam-Rijnmond*. The system will structure, analyze, and visualize large amounts of data that are generated (partly in real-time), such as camera pictures of drones or body cams, characteristics of incidents, building properties, historic incident data, etc. This introduction of AI technology offers new possibilities for more efficient and effective decision-making and operations, e.g., prediction of the development of an incident or advice on decisions based on previous incidents; however, the design and development should be done responsibly, taking human values of relevant stakeholders into account.

Research questions in the project concern the identification of expectations of relevant stakeholders, consequences for the organizational structure, how the introduction of AIDA influences work processes and information processing, and which values and value tensions occur when using AIDA in future scenarios. In this research, to identify the impact on human values of relevant stakeholders of the system, we conducted 10 focus group sessions with fire service personnel in which we assess and analyze the stakeholders' values. This article discusses the design and implementation of the study and, as we are still in process of analyzing the sessions in detail, summarizes preliminary insights.

In the following, we first present background on human values for AI development and give an overview of the envisioned AI system AIDA. We then describe the design and implementation of the study and preliminary results, followed by an outlook on next steps and discussion of the results.

BACKGROUND

In this section, we describe the importance of including human values into AI development for FR, give a short overview of previous research on the identification of affected human values, and describe the envisioned decision-support system AIDA, its functionalities and how it will be integrated into the current situation.

Including Human Values into AI Development for FR

In the last years, a lot of attention has gone towards the responsible use and development of AI, amongst others by the EU High-level expert group on AI (AI HLEG, 2019) and IEEE (IEEE, 2021). The AI HLEG identified 4 ethical principles that AI systems should adhere to (i.e., respect for human autonomy, prevention of harm, fairness, and explicability) and 7 key requirements (i.e., human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, societal and environmental well-being, and accountability). The AI HLEG stresses that these ethical principles and requirements need to be situated and considered in context and that it should be determined for the specific applications whether these requirements are applicable, possibly supplementing them with additional relevant requirements. This means that for all AI applications, the ethical aspects need to be investigated to determine relevancy and applicability. IEEE also emphasizes the importance of considering ethical values throughout the stages of concept exploration and development.

The importance of including human values into the design and development of (AI) systems is also at the basis of the Value-sensitive design methodology (VSD) (Friedman et al., 2013). VSD is an iterative methodology that integrates three perspectives, namely conceptual, empirical, and technical investigations. These investigations are executed iteratively, though not necessarily in a set order. In this research, we focus on the empirical investigations for identifying impact of AIDA on fire services.

In previous work, ethical issues have been investigated for example for Search and Rescue robotics. Harbers et al. (Harbers et al., 2017) identified and analyzed relevant values affected by rescue robotics; values that were identified were, amongst others, personal safety, safety of others, access to information, well-being, effectiveness, ease of use, authority, health, and contact. Battistuzzi et al. (Battistuzzi et al., 2021) conducted a scoping review on ethical concerns in rescue robotics, identifying seven core ethically relevant themes: fairness and discrimination; false or excessive expectations; labor replacement; privacy; responsibility; safety; trust. In Mioch et al. (2024), as a

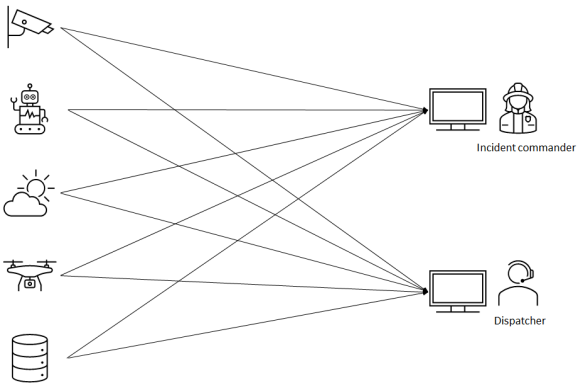

Figure 1. Current system set-up.

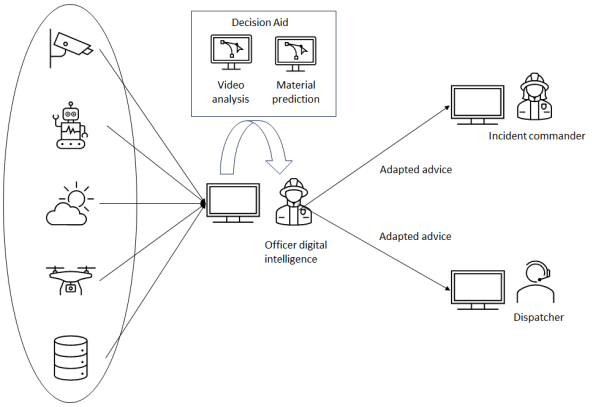

Figure 2. Envisioned set-up with AIDA.

step towards a generic framework of ethical aspects for AI in FR, relevant direct and indirect stakeholders of AI systems for fire services were identified and relevant human values of a prioritized selection of these stakeholders regarding the application of AI systems were investigated. The following stakeholders were identified as most relevant to take into account during the introduction of AI systems: fire fighters, incident commanders, special operations fire fighters (e.g., for hazardous materials and digital exploration), dispatchers, company doctors, incident researchers, citizens, and the surroundings (being an umbrella term for the nearby environment, including objects, people, nature, etc.). Positive as well as negative impact was identified for different human values. Starting point of the identification of values was the account of Friedman et al. (Friedman et al., 2017) as these are considered to be particularly important for technological design. Several values as defined by Friedman et al. were found to be impacted (mostly) positively for the different stakeholders (although there was no value that is only affected positively), i.e., *physical well-being* and *psychological well-being*. Several values were found to be impacted only negatively for the different stakeholders, i.e., *autonomy*, *identity*, *informed consent*, *privacy*, and *trust*. Several other values were identified to be impacted by AI systems, i.e., *accountability*, *transparency*, *reliability*, *security*, and *appropriate training*. AI systems should (explicitly) support the possible positive impact that they could have whereas the negative impact should either be reduced or avoided altogether, e.g., by specifying design requirements that explicitly take these human values into account. In this research, we will further investigate the impact on human values of the (envisioned) AI system AIDA, situating and applying the research by Mioch et al. (2024) for this specific application to further determine relevance of found values.

AIDA

A lot of different data is being generated or accessible by fire services, such as camera images of drones or body cams and building properties; currently, this data is streamed directly to FRs without any analysis or interpretation and displayed on different (sub)screens, see Figure 1. That means that FRs continuously need to monitor large amounts of data to build good situation awareness (SA). This is very time– and work-intensive and can lead to cognitive overload of individual FRs and subsequently to suboptimal decision-making, especially during complex and dynamic incidents and high time pressure. To mitigate these effects, the Dutch Safety Region *Rotterdam-Rijnmond* is developing an AI system, AIDA, to support them in the analysis, interpretation, and visualization of this (real-time) data. This introduction of AI technology offers new possibilities for more efficient and effective decision-making and operations, e.g., prediction of the development of an incident based on data from comparable incidents and information on the current status. In addition, new roles are envisioned to support in the sense-making process and sharing of information.

For AIDA, different modules are being developed. For this research, we will focus on two modules that are designed to help the dispatcher and the incident commander with building up situation awareness and their decision-making, i.e., the *video analysis module* and the *material prediction module*. The main goal of the former is to support shared situation awareness, whereas the latter will provide advice on the material to send to an incident. For both modules, AI systems are envisioned. Shared situation awareness (of for example the incident commander) in the *video analysis module* will be supported by analysing and interpreting different video sources (e.g., drone camera pictures) to identify relevant information, such as locating humans in camera or drone pictures in the incident area faster and with less cognitive effort, aiding with decision-making regarding the deployment of teams and logistical effort (a first prototype description can be found in Heemskerk et al. (2024)). This will be particularly useful in situations in which the situation is very dynamic and complex. Advice in the *material prediction module* will

Table 1. Overview over scenario selection per session and participants.

Session	*Scenario*	*# (Head) Incident Commanders*	*# Dispatchers*	*# Fire fighters*	*# Emergency co-ordinators*
1	(Adapted) Scenario 2	2	2		
2	Scenario 2	1	1		1
3	Scenario 1	2	1		
4	Scenario 2	2	2	1	
5	Scenario 1	1	1		
6	Scenario 1	2	2		
7	Scenario 2	2	2	1	
8	Scenario 1	2	2		
9	Scenario 2	1	3	1	
10	Scenario 1	2	3		
Total		*17*	*19*	*3*	*1*

be given on, amongst others, the expected material that is needed during particular incidents, based on data of previous incidents, real-time characteristics of incidents and other data such as cadastre data. The safety region expects to introduce a new role, the *Officer Digital Intelligence (ODI)*, who monitors and analyses the different data streams and outputs of the different modules and coordinates the information towards the dispatcher and incident commander. Both keep access to all data, but are supported by the ODI, see Figure 2. In the following, the two modules are described in more detail.

Video Analysis

A lot of camera pictures are generated by different cameras during incidents. One source of video data are drones that film the location of an incident from above. The video feeds need to be monitored to recognize relevant information for incident response, such as development of the fire or presence of people. The envisioned AI module will analyze and interpret the video stream and detect humans. It will be able to classify humans into fire fighters, ambulance personnel, and civilians. In addition, a short summary can be made of the video stream based on changes and development of presence of people in the footage.

Material Prediction

A lot of data is available from past incidents, as the development of each incident is logged, e.g., the address of an incident, characteristics of the incident, and material that was requested (type of vehicle, specialization of personnel, etc.). The envisioned AI system will, based on correlations found in the available past data, predict what kind of material is needed for an incident and advise the officer to request the necessary material.

METHOD

To determine the relevant values of different stakeholders for AIDA, we conducted 10 focus group sessions with different stakeholders of the fire services. In total, 40 employees of the fire services participated, of which 1 head incident commander, 16 incident commanders, 19 dispatchers, 3 fire fighters, and 1 emergency coordinator. Many of the participants had several roles in the fire services, e.g., a dispatcher also working as voluntary fire fighter; for this overview, we selected the role that the participants mentioned when signing up. 5 participants were female. For an overview of the participants, see Table 1.

During the focus group sessions, the following steps were followed: first, a short introduction into AI technology was given, its working, possibilities, and limitations, to make sure that participants had a general knowledge of the working of AI systems to be able to identify impact of AI. Most participants had heard of the plans to introduce AIDA, but were not familiar with details. For that reason, the envisioned functionalities of AIDA were introduced as described above, together with the envisioned embedding into the organisation. Two scenarios, one for each module, were used to instantiate the working of AIDA into situation and organizational processes, see Table 2. These scenarios highlighted value tensions. In each session, one of these scenarios was presented as a basis for discussion on the positive and negative impact of the corresponding AIDA module. As a first step for this research, the participants were asked to identify positive and negative impact of AIDA for the roles *Incident commander*, *Dispatcher*, and *Officer Digital Intelligence*. These roles were chosen because (most) participants belonged to at least one of these stakeholder groups, these stakeholder groups will be directly influenced by the results of the

Table 2. Scenario descriptions.

	Scenario 1 - Video analysis module	*Scenario 2 - Material prediction module*
Situation	Large fire in industrial building; The fire brigade explores the situation. A drone team is called and a drone provides a live camera feed of the building and environment. It becomes clear that not all employees are accounted for.	A building in the center of Rotterdam is on fire. The dispatcher receives the 112 call with the location of the fire and notifies the fire brigade with the information; according to procedure, two cars are sent.
Data	Drone camera stream; camera pictures of bodycams.	Historical incident data, characteristics of current incident.
AI technology	The AI system processes the data in real-time and combines the various data sources; it detects humans in the camera pictures and distinguishes between fire fighters, ambulance personnel, and civilians. The system can also give a summary of the video feed.	The AI system predicts, based on historical incident data, what kind of material will be needed. In an adapted version (session 1), the AI system predicts the personnel that should be called up to be deployed (instead of which material should be deployed).
Steps	The ODI requests a summary of the last minutes of video stream, in which newly recognized persons are displayed in the video stream. AIDA identifies people on the video stream and shows a summary of all people found. The system distinguishes between firefighters, ambulance personnel, and civilians. The ODI looks at the summary of the video stream. Several citizens are shown. The ODI provides the relevant information to the dispatcher.	AIDA continuously monitors and assesses the material requirements of incidents, based on various data sources such as land registry data and historical incident data. It predicts that only one car is needed for this incident and passes this on to the ODI. The ODI assesses the advice, agrees with it and passes it on to the operator.

system, and because these stakeholder groups are amongst the stakeholders that were identified in previous research as particularly relevant (Mioch et al., 2024). In the continuation of this research, impact on other stakeholders will also be investigated. The results were shared plenary and discussed. During the discussion, additional impact was added. Also, the participants were asked to discuss how positive impact could be supported and negative impact be reduced. During the focus group sessions, we tried to focus on the impact of the AI modules and not on the impact or the introduction of the new role (the *ODI*).

PRELIMINARY RESULTS

As we are still in the process of analyzing the expert group sessions in detail, in this section, we will describe preliminary results, i.e., a selection of expected impact that was mentioned by the participants. First, we shortly describe impact and requirements of the new role that is suggested to support dispatcher and incident commander with the analysis of the data. After, for each of the two AI modules, we describe a selection of impact participants mentioned and corresponding human values that are implicated, based on the previously identified list of human values that are implicated by AI systems for FR (Mioch et al., 2024). As we are still in the process of analysing the results, these first impressions will later be updated.

New Role *Officer Digital Intelligence* As mentioned above, the introduction of AIDA is accompanied with the introduction of a new role, the *ODI*. Various aspects were mentioned in the sessions that are relevant to the role of ODI, and especially to the collaboration with the incident officer and the dispatcher. First of all, the expertise of the ODI is very important; the ODI needs to have good insight into the tasks and role of both dispatcher and incident commander. Since the dispatcher and officer can be very busy with their own tasks, it is very important that the ODI understands what kind of information the other two need or do not need at any time, so as not to disrupt their process, introduce overload, or miscommunication (by speaking 'two languages'). The ODI should take work off the hands of the dispatcher, but should not make decisions; the ODI should have a support function and work together with the dispatcher and the incident commander.

Video Analysis

The participants discussed the impact of the *Video analysis module* in 5 sessions. In the following, we present a selection of expected impact that was mentioned by the participants, sorted by value or ethical aspect that is affected. Please note that this overview is by no means complete yet and only gives an idea on some of the topics that came up.

Autonomy Autonomy refers to people's ability to decide, plan, and act in ways that they believe will help them to achieve their goals (Friedman et al., 2013). Impact that affects *autonomy* was mentioned for the introduction of both modules. Regarding the module *Video analysis*, participants stated that the systems might lead to less decision-making power because of valuing and interpreting information becomes more difficult, e.g., because of not understanding the information and analysis that the system provides ("But the problem with the system is that it might be better than we are, and that we cannot understand it. [...] It is very difficult to find a good balance: to some extent you want to use it, but otherwise you also want to handle the situation yourself."). Participants mentioned that they are concerned about their dependence on the technology and that if the system fails, their ability to operate might be impaired because they forgot how to handle the situation without the system. Specifically for the application of detecting and summarizing persons in the drone video stream, participants were concerned that it would influence autonomy negatively through disproportional trust in the results of the system ("It's exactly like that video where the monkey walks through the screen where 100 people don't see that monkey, [...], you no longer see the rest of the information, because I become triggered to look at people.")

Identity Identity refers to people's understanding of who they are over time, embracing both continuity and discontinuity over time (Friedman et al., 2013). Impact that related to the value *Identity* was mentioned in several sessions, for example regarding expected changes in tasks for the different stakeholders, e.g., positive impact on the dispatcher ("the dispatcher can focus on his or her main tasks"), but also the expectation that some tasks will no longer be performed by the dispatcher ("The ODI takes quite a bit of work off the dispatcher's hands"). Some participants were afraid that their role will become less relevant and that they will be skipped in the communication between ODI and incident commander. The dispatcher's expertise could be used less which led to feeling threatened in their work ("That can [...] be kind of demotivating for the dispatcher or feel as if they become redundant or irrelevant, or less important").

Psychological well-being Psychological well-being falls under the value *Welfare* as identified by Friedman et al. (2013). Participants mentioned that introducing (summaries) of video streams of incidents for the ODI or the dispatcher will give them more information and overview on what is happening and might improve decision-making. However, it will also introduce pictures of what is happening, which might lead to more PTSD (" you see something that you may not want to see in that role").

Trust in AI Participants mentioned that (appropriate) trust in AI systems is important. Participants said that there is a danger of too much trust, e.g., being influenced in decision-making too much, not evaluating and reflecting enough the situation ("I'm a bit afraid of tunnel vision; that we will focus 100 percent on the system [...] and that common sense will fade into the background"). In addition, it is very important to understand the capabilities and limitations of the system to be able to determine how to include the pictures in the decision-making process ("[...] that you are going to make assumptions, that you do not check whether the pictures from the drones are actually current reality, or that there are aspects that are not filmed.").

Accountability Accountability can be seen in the context of liability (legal responsibility) as well as social responsibility. Participants mentioned that they expect negative impact on accountability, for example when the system provides information that is not followed on by the FRs ("And I think you will soon have another problem, that the information is there, [...] but you have not done anything with it. And then you get public pressure [...] and you will get blamed for it.").

Material Prediction

The participant discussed the impact of the *Material prediction module* in 4 sessions. In the following, we present a selection of expected impact that was mentioned by the participants, sorted by value or ethical aspect that is impacted.

Autonomy Regarding *Autonomy*, participants mentioned that it might be more difficult for dispatchers to make decisions, as it might be difficult to interpret the advice and integrate it in the decision-making process ("We now always say the caller is always right, until proven otherwise. So if the caller says there is a raging fire and the system says that based on that criteria, you have to scale up or down, how do I deal with that?"). In addition, they were concerned about depending too much on the system and about their ability to operate when the system fails ("Are you still sharp at the moment the system does something crazy?").

Trust in AI Participants mentioned that they have difficulty in trusting an AI system that gives advice on aspects that might lead to a less optimal task performance, for example, in the case of sending less material to an incident ("I feel anxious (apprehensive) about scaling down based on technology."). In addition, even if they could understand the advice and rationally, this advice seems sound, then these kind of decisions still also have an emotional aspect that makes it difficult to follow the advice ("You may be able to explain it all rationally, but your feeling... says something different."). Also, they mentioned that there always is a difference between a theoretical reality (e.g., houses that are optimally built regarding fire safety, but where the context such as changes in the building or the inhabitants counteract the optimal circumstances) and practice, which makes systems based on data inherently unreliable.

Accountability Participants are concerned that legal responsibility need to be clearly specified; for example, when following advice of the AI system although it is not correct ("If I scale down and it is nevertheless serious, then it's my fault"). In addition, participants feel a strong social responsibility and see it critically that the *Material prediction* could advise less material to be dispatched ("It goes against your sense of helping").

Privacy Privacy refers to a "claim, an entitlement, or a right of an individual to determine what information about himself or herself can be communicated to others" (Friedman et al., 2013). Impact on the value *privacy* was only mentioned during the first session, in which the scenario (and AI module) was adapted towards planning of which personnel should be called up to be deployed (instead of which material should be deployed). During this session, the topic was discussed fervently, as this directly had links to the monitoring of personnel. Impact on privacy for citizens was not mentioned during the sessions. This might be because the participants were not explicitly asked to identify impact for these stakeholders.

DISCUSSION

Previous research (Mioch et al., 2024) investigated which human values are impacted by AI systems for fire services in general; the investigation was set-up to identify these values broadly and the impact of AI systems was not contextualized. This resulted in an overview of values that might be impacted (positively and negatively), but not in concrete (design) requirements on how to support positive impact and decrease negative impact. With this research, we contextualize this previous research by choosing a concrete AI application and discussing the impact of two example modules on fire services practice. By eliciting expected impact for these concrete AI modules, translating this impact towards affected human values and linking design requirements to these values, we hope to be able to make the step towards defining grounded design requirements for the responsible design and development of these modules.

In this paper, we have made first steps towards identifying relevant human values for two modules of AIDA. To achieve this, we have held 10 focus group sessions, 5 for each module, in which we discussed impact of the modules on tasks and roles of the FRs; we take a socio-technical systems approach, expecting changed task allocations and -responsibility and new human-AI dependencies. We presented first results of these sessions, i.e., statements that can be linked to value considerations, which again can influence design requirements. These design requirements should (explicitly) support the possible positive impact on relevant values whereas negative impact should either be reduced or avoided altogether, e.g., by specifying design requirements that explicitly take these human values into account. There are several things that can be noted on the basis of the preliminary results: first of all, several of the previously identified human values are implicated by the AI modules, e.g., *Autonomy* and *Identity*. In addition, values such as *Trust in AI* and *Accountability* have been mentioned for both modules. During the focus group sessions, the participants discussed the impact of the two AI modules within the context of two concrete scenarios and their fire services practice, including obstacles they see for the introduction of these specific modules and possible implementation choices to enhance positive impact; this discussion will be the basis for the analysis of the design requirements.

During the sessions, it was difficult to keep discussion to the specific scenarios and impact of the selected AI module. Possibly, the scenarios we presented were not specific enough and, in the future, should be made more specific by a simulated collaboration environment (as used in e.g., (Beuker et al., 2021)). Also, possibly because of the very different (technical) expertise and background of the participants, the discussions not always focused on the expected AI-related impact, but also on having access to particular data or other impact of IT-related systems currently in use.

We realize that, as described above, the introduction of AIDA is very complex and that it is not (yet) possible to discuss impact on the tasks, processes, and responsibilities of the different roles, as this also includes organisational changes through the introduction of a new role, the *ODI*. We only superficially touched on this issue when describing preliminary requirements on the expertise of the ODI. In the future, we will also look into these organizational aspects in more detail and analyse, discuss, and advise on the organizational and procedural changes. We also realize that for the participants, it was a challenge to differentiate between the impact of AIDA and the impact of the organizational change. We will take this into account in the analysis of the results.

The focus group sessions have been held right at the beginning of the modelling and development process for AIDA. The functionalities of the system are not yet set. The modelling and development process will profit from our research in several ways: (1) based on positive and negative impact that has been identified, design requirements will be specified that support the positive impact and reduce negative impact, (2) explicitly, values of different stakeholders are taken into account during the design and development process, and (3) the different stakeholders will be regularly involved in the research, which facilitates support and acceptance and improves the quality and suitability of the system. For now, we focused this effort of identifying impact of the two AIDA modules on three stakeholder groups from the fire services (i.e., *dispatcher*, *incident commander*, and *officer digital intelligence*); in the future, we would also like to hold focus group sessions with other FR organizations (e.g., police, ambulance services) and citizens to also take their perspectives into account.

In the future, we will further analyse the results of the expert group sessions and plan to (1) compare value implications between the two modules ('Are different values affected by the two modules?') and (2) compare identified values with the general framework ('Are there additional values affected that have not been identified before, or are some previously identified values not affected by these two modules?'). We will then integrate the results into the ethical framework, together with values found from normative sources, adding results of a scoping literature review on ethical aspects of AI systems applied to first response. The ultimate goal is to develop a framework for responsible development and use of AI in the fire services domain, including requirements and methods (e.g., with regard to identifying values and value tensions), models (e.g., with regard to situated values), design patterns of responsible AI development and deployment, and evaluation methods for responsible AI so that practitioners and AI developers can use the framework as a toolbox towards responsible AI design, development, and deployment.

ACKNOWLEDGMENTS

We would like to thank the first responders of *Gezamenlijke Brandweer Rotterdam*, *Veiligheidsregio Rotterdam-Rijnmond*, and *Veiligheidsregio Zuid-Holland-Zuid* for their participation in this research. This research was supported by the University of Applied Sciences Utrecht (HU) through a '*promotievoucher*'.

REFERENCES

AI HLEG. (2019). Ethics Guidelines for Trustworthy AI. European Commission.

Battistuzzi, L., Recchiuto, C., & Sgorbissa, A. (2021). Ethical concerns in rescue robotics: A scoping review. *Ethics and Information Technology*, *23*(4), 863–875. https://doi.org/10.1007/s10676-021-09603-0

Beuker, T., Mioch, T., & Neerincx, M. A. (2021). Team Design Patterns for participatory development of First Response Human-Agent Teaming. In W. Karwowski, T. Ahram, M. Milicevic, D. Etinger, & K. Zubrinic (Eds.), *Human systems engineering and design (IHSED 2021): Future trends and applications. AHFE (2021) international conference* (Vol. 21). AHFE Open Access. https://doi.org/10.54941/ahfe1001151

ECP. (2019, January). *Artificial Intelligence Impact Assessment (English version)*. Retrieved February 5, 2024, from https://ecp.nl/publicatie/artificial-intelligence-impact-assessment-english-version/

European Commission. (2021). *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts.* (final No. COM/2021/206).

Friedman, B., & Hendry, D. G. (2019). *Value sensitive design: Shaping technology with moral imagination*. Mit Press.

Friedman, B., Hendry, D. G., & Borning, A. (2017). A Survey of Value Sensitive Design Methods. *Foundations and Trends® in Human–Computer Interaction*, *11*(2), 63–125. https://doi.org/10.1561/1100000015

Friedman, B., Kahn, P. H., Borning, A., & Huldtgren, A. (2013). Value Sensitive Design and Information Systems. In N. Doorn, D. Schuurbiers, I. van de Poel, & M. E. Gorman (Eds.), *Early engagement and new technologies: Opening up the laboratory* (pp. 55–95). Springer Netherlands. https://doi.org/10.1007/978-94-007-7844-3_4

Harbers, M., de Greeff, J., Kruijff-Korbayová, I., Neerincx, M., & Hindriks, K. (2017). Exploring the ethical landscape of robot-assisted Search and Rescue. *Intelligent Systems, Control and Automation: Science and Engineering*, *84*, 93–107. https://doi.org/10.1007/978-3-319-46667-5_7

Heemskerk, J., Mioch, T., Maathuis, H., & Aldewereld, H. (2024). Long-range human detection for first response decision support. In B. Penkert, B. Hellingrath, M. Rode, A. Widera, M. Middelhoff, K. Boersma, & M. Kalthöner (Eds.), *Proceedings of the 21st ISCRAM conference*.

IEEE. (2021). IEEE standard model process for addressing ethical concerns during system design. *IEEE Std 7000-2021*, 1–82. https://doi.org/10.1109/IEEESTD.2021.9536679

Martinez-Alpiste, I., Golcarenarenji, G., Wang, Q., & Alcaraz-Calero, J. M. (2021). Search and rescue operation using uavs: A case study. *Expert Systems with Applications*, *178*, 114937.

Meina, M., Ratajczak, E., Sadowska, M., Rykaczewski, K., Dreszer, J., Balaj, B., Biedugnis, S., Węgrzyński, W., & Krasuski, A. (2020). Heart rate variability and accelerometry as classification tools for monitoring perceived stress levels—a pilot study on firefighters. *Sensors*, *20*(10). https://doi.org/10.3390/s20102834

Mioch, T., Aldewereld, H., & Neerincx, M. A. (2024). Empirical investigation of values affected by AI systems for fire services. In B. Penkert, B. Hellingrath, M. Rode, A. Widera, M. Middelhoff, K. Boersma, & M. Kalthöner (Eds.), *Proceedings of the 21st ISCRAM conference*.

Stahl, B. C. (2021). Ethical Issues of AI. In B. C. Stahl (Ed.), *Artificial Intelligence for a Better Future: An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies* (pp. 35–53). Springer International Publishing. https://doi.org/10.1007/978-3-030-69978-9_4

Talaat, F. M., & ZainEldin, H. (2023). An improved fire detection approach based on yolo-v8 for smart cities. *Neural Computing and Applications*, *35*(28), 20939–20954.