Empirical Investigation of Values Affected by AI Systems for Fire Services

Tina Mioch*
Delft University of Technology
University of Applied Sciences Utrecht
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

Fire fighters operate in a dangerous, dynamic, and complex environment. Artificial Intelligence (AI) systems can contribute to improve fire fighters' situation awareness and decision-making. However, the introduction of AI systems needs to be done responsibly, taking (human) values into account, especially as the situation in which fire fighters operate is uncertain and decisions have a big impact. In this research, we investigate values that are affected by the introduction of AI systems for fire services by conducting several semi-structured focus group sessions with (operational) fire service personnel. The focus group outcomes are qualitatively analyzed and key values are identified and discussed. This research is a first step in an iterative process towards a generic framework of ethical aspects for the introduction of AI systems in first response, which will give insight into the relevant ethical aspects to take into account when developing AI systems for first responders.

Keywords

Values, Fire services, Value sensitive design, Responsible AI.

INTRODUCTION

First responders (FRs) operate in a dangerous, dynamic, and complex environment, in which they have to quickly understand the situation and how to mitigate danger, while keeping themselves and civilians safe. To improve effectiveness and safety, FR needs and corresponding capability gaps were identified by the International Forum to Advance First Responder Innovations (IFAFRI), e.g., the ability to create actionable intelligence based on data and information from multiple sources (Capability Gap 9) and the ability to conduct on-scene operations remotely without endangering responders (Capability Gap 7) (IFAFRI, 2019). Technology could enhance the safety, effectiveness, and efficiency of FRs and help closing these capability gaps.

One of the technologies which has a high potential for contributing to filling these capability gaps is artificial intelligence (AI). AI systems can play an important role in the future to, for example, improve shared situation awareness (e.g., (Mioch et al., 2021)) and contribute to advanced decision-making (Radianti et al., 2019). This leaves FR organizations with the challenge of integrating these AI systems in the decision-making processes in a responsible way by addressing (public) values and public function properly. In our view, AI systems in this domain are always part of a hybrid human-AI system, a socio-technical system, in which task allocation and task responsibility might change and new human-AI dependencies arise. This introduces the research challenge of developing hybrid human-AI systems in which AI technology and humans cooperate in a way that synergy is created (Akata et al., 2020; Seeber et al., 2020). In such a hybrid intelligence system, complex goals can be accomplished by combining

*corresponding author

humans and AI technology that collectively work on shared objectives with complementary capabilities that, when combined, augment each other (Dellermann et al., 2019).

The development and application of these hybrid human-AI systems need to be done responsibly (AI HLEG, 2018), e.g., regarding possible biases in (training) data sets and privacy aspects (e.g., (personal) data on FRs such as location, performance, stress). For some domains, the ethical aspects of AI systems and applications have received a lot of attention, e.g., the health domain (Blasimme & Vayena, 2020; Morley et al., 2020; Murphy et al., 2021) and the military domain (Galliott & Scholz, 2020; Wasilow & Thorpe, 2019). However, in the field of FR, it seems that ethics with regard to the application of AI has not yet been much addressed. To develop and apply AI systems responsibly, it is important that the AI systems support the stakeholders' values and that values of different relevant stakeholders are taken into account. The context in which an AI system is applied and choices in the (technical) development of the AI system determine the relevance and importance of the different values. In previous work in the FR domain, key ethical concerns have been identified for Search and Rescue robots to support development and deployment in a responsible way (e.g., see Harbers et al. (2017) for empirical research and Battistuzzi et al. (2021) for a scoping review). Following on this work, in this paper, we provide a qualitative empirical investigation of relevant values of relevant stakeholders of AI systems for fire services. This is a first step towards a generic framework of ethical aspects for AI in FR, which will give insight into the relevant ethical aspects to take into account when developing AI systems for FR. Subsequent steps for setting up the generic framework are differentiating between general values and instances of these values in specific use cases to take the specific context into account and linking and updating the framework with societal developments regarding ethical aspects such as norms and guidelines.

This research is inspired by the Value Sensitive Design (VSD) methodology, which accounts for human values throughout the design process (Friedman & Hendry, 2019). First, we identify relevant direct and indirect stakeholders of fire services. Second, we investigate relevant values of these stakeholders regarding the application of AI systems. To do this, we conducted three focus group sessions with fire service personnel in which we assess and analyze the stakeholders and their values. Using the focus group session results, we identify a first set of key values to take into account in the application of AI systems for fire services.

In the following, we first give background on ethical themes for AI systems and on values and value-sensitive design. We then describe the setup and execution of our focus group sessions and the results of these sessions, followed by the conclusion and discussion of the results.

BACKGROUND

In the last few years, a lot of attention has gone towards the responsible development of AI, amongst others by the EU High-level expert group on AI (AI HLEG, 2018)) and IEEE (IEEE, 2021). The AI HLEG was appointed to advise on a European AI strategy and identified key requirements, which each AI technology needs to fulfil before it is considered safe and trustworthy (AI HLEG, 2018). According to these guidelines, trustworthy AI systems should be lawful (i.e., respecting all applicable laws and regulations), ethical (i.e., respecting ethical principles and values), and robust (i.e., both from a technical perspective while taking into account its social environment). Four ethical principles are identified that AI systems should adhere to, namely respect for human autonomy, prevention of harm, fairness, 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 wellbeing, and accountability. These requirements are explicitly non-exhaustive. When applying these guidelines to specific AI applications, the ethical principles need to (1) be situated and considered in context to determine their relevance and supplement them with possible additional ethical requirements and (2) be translated into design requirements for the specific application and context. The AI HLEG does not give guidelines on how to do this and they have not been applied structurally to the FR domain.

Technology is not neutral but always manifests values (van de Poel, 2015). Values refer to what a person or group of people considers important in life (Friedman et al., 2013). Schwartz (2012) identified 10 basic universal values, i.e., power, achievement, hedonism, stimulation, self-direction, universalism, benevolence, tradition, conformity, and security. Some of these values conflict with each other (e.g., benevolence and power), whereas some are compatible with each other (e.g., conformity and security). A possible approach to support taking values into account in technology design is Value sensitive design (VSD). VSD is "a theoretically grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner throughout the design process" (Friedman et al., 2013, p. 56). Values that are considered particularly important for technological design are Welfare, Ownership and Property, Privacy, Freedom from Bias, Universal Usability, Trust, Autonomy, Informed Consent, Accountability, Courtesy, Identity, Calmness and Environmental Sustainability. As we are

investigating values affected by AI systems for fire services and not general human values of stakeholders within the fire services, we use Friedman et al.'s account of values as a starting point in our investigation.

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. VSD contains a rich collection of different methods that help designers to investigate values in technology, such as direct and indirect stakeholder analysis and value scenarios (Friedman et al., 2017), and is thus particularly useful for our research. To be able to determine human values that should be taken into account, VSD asks system designers to establish a robust set of stakeholder groups and to justify those who likely are most strongly affected – that is, to provide an conceptual or empirical rationale for their inclusion in the design process. The same holds for values; the applicability and relevance of values should be investigated from the three different perspectives. In this research, we focus on one of VSD's perspectives, namely empirical investigation, by qualitatively identifying a set of stakeholders to include and key values to take into account when developing AI for fire services. In later steps, we will iteratively extend, refine, and test the results, also from a conceptual and technical perspective.

A tool that can help with creating an overview of the potential positive and negative impact of AI systems is the *Ethical matrix* (Mepham et al., 2003). The Ethical matrix is a conceptual tool designed to help decision makers deliberate about the ethical acceptability of existing or envisioned technology. It makes the impact of design choices on the different stakeholders explicit and provides structure and support in the design process. The cells of the matrix contain the impact, negative or positive, of the envisioned technology, on each of the stakeholder groups for specific ethical values.

METHOD

To determine relevant values for fire services regarding the application of AI, we conducted semi-structured interviews during focus group sessions with different subject-matter experts from the fire services. The goal of the focus group sessions was exploratory, to gain as much information as possible on expected impact of AI systems for fire services, and the results were qualitatively analysed.

The first focus group was conducted as an exploratory session with three incident commanders that are involved in innovation projects. The goal was twofold, first, to start exploring relevant stakeholders and expected impact of AI systems on these stakeholders, second, to test and evaluate the focus group setup. Two other focus group sessions with each 4 participants were conducted with fire fighters (carrying out day-to-day firefighting and fire safety work; 3 participants), dispatchers (managing emergency calls, ensuring that proper response teams are sent to the incident location; 2 participants), incident commanders (in charge of an incident; 2 participants), and an HR professional (1 participant). All participants (besides the HR professional) had experience in the field, ranging from limited experience (1-3 years experience) to very experienced (15+ years of experience). The focus group sessions lasted 2 hours.

The first, exploratory, focus group session consisted of several steps. To identify as many stakeholders as possible, the participants were asked to list all people that interact with an envisioned AI system (presented in a general scenario) and people that are affected by the system or who have a vested interest in its success or failure. To not have any influence effects on the process, the experts were asked to do this by themselves. After 5 minutes, the results were shared and discussed. Based on the combined list of stakeholders, a prioritization of stakeholders was made. After having selected the most relevant stakeholder, the participants determined which positive or negative impact of AI systems they could identify for these stakeholders. The participants went through the stakeholders (by themselves) and wrote identified impact on post-its. The results were plenary shared and discussed; as every participant shared their identified impact, participation of each of them was ensured. Additional impact was added during the discussion. After the focus group session, the identified (positive and negative) impact was analyzed regarding (possible) underlying values by a human-AI collaboration expert, mapped on an ethical matrix (van der Stappen & Steenbergen, 2020) and presented to the participants for validation. The discussion of this first version of the ethical matrix led to additional input regarding impact for stakeholders.

The other two focus group sessions consisted of the same steps as the first exploratory focus group, with slight adaptations, based on lessons-learned of the first focus group session: we realized that (1) a general knowledge of the working of AI systems is needed to identify impact of the systems and (2) example scenarios support the identification of impact. The participants of the first focus group were all involved in (AI) innovation projects and familiar with the possibilities of AI technology; for the other two focus groups, this was not the case. For that reason, we added a short overview over AI technology, its working, possibilities, and limitations. Also, two concrete scenarios (inspired by examples mentioned during the exploratory focus group session) were introduced as basis for the discussion on positive and negative impact of AI systems. These (descriptive) scenarios highlighted value

Table 1. Scenarios used during the second and third focus group.

	Scenario 1	*Scenario 2*
Situation	Large fire in storage tank; Fire brigade team is exploring the situation. Fire fighters wear sensors.	Tank truck with hazardous material collides with a tree within city limits. Hazardous substances spill.
Data	Measurement of physical properties of firefighters such as stress level, position, body temperature, presence of dangerous gases, body cam.	Video data from road cameras is processed automatically. Data from past incidents is available.
AI technology (first scenario version)	An AI system processes the data in real-time and combines the various data sources; it provides information to the incident commander about the current status of the firefighters in the field, monitors the situation and stress, and warns the incident commander in case of conspicuous things, e.g., if fire fighters come too close to a danger zone.	An AI system reports the incident to the dispatcher; it recognizes hazardous substances through ADR sign recognition and passes on this information. It also predicts the development of the situation based on similar incidents in the past and recommends to the dispatcher which resources should be sent. The dispatcher forwards reports and information to the fire house.
AI technology (second scenario version, extended autonomy of AI system)	Continuously monitors and predicts the (stress) status of the firefighters, also by means of stress-related data from the past. System advises the incident commander on employability (short-term and long-term)	The AI system reports the incident to the dispatcher, assesses who needs to be called, and puts the call through to the responsible fire house. It recognizes hazardous substances through ADR sign recognition and also passes on this information. It predicts situation development based on similar incidents in the past and forwards call for appropriate resources. The dispatcher monitors the report.

tensions. We described our scenarios as 'what if' scenarios, scenarios that are situated just a few years into the future and inviting the participants to assess ethical issues posed in the scenario (Wright et al., 2014), see Table 1. Each scenario consisted of two versions, with the second version extending the AI's autonomy to increase (possible) ethical issues.

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. We categorized the resulting 28 codes into 13 themes, which will be described in the next section.

RESULTS

In this section, we describe the results of the focus group sessions. First, an overview is given of the most important stakeholder groups that were identified by the participants, together with the human values (as identified by Friedman et al. (2013) that are affected by the identified impact. Subsequently, the identified impact on the human values is described in more detail and in context of fire services, followed by other, more AI-related values that participants mentioned regarding the introduction of AI systems for fire services.

Stakeholders and Values

During the focus groups, participants identified the following stakeholders 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. For these stakeholders, the participants identified possible positive and negative impact of AI applications. Most impact was identified for three stakeholder groups, namely fire fighters, incident commanders, and dispatchers. This is not surprising, as the scenarios involved these stakeholder groups explicitly, and these stakeholder groups were (mostly) represented by the participants during the focus groups. Table 2 provides an overview of the (most relevant) stakeholders as identified during the focus group sessions together with the human values (as specified by Friedman et al. (2013)) which the identified impact affects.

Table 2. Most relevant stakeholders as identified during the focus group sessions and values (as specified by Friedman et al. (2013)) that the identified impact affects.

Stakeholders	*Values*
Incident commanders	Autonomy, identity, physical well-being
Fire fighters	Autonomy, identity, informed consent, privacy, physical well-being, psychological well-being, trust
Dispatchers	Autonomy, identity, privacy, physical well-being, psychological well-being
Citizens	Privacy, physical well-being, psychological well-being, trust
Surroundings	Physical well-being
Special operations fire fighters	Physical well-being
Incident researchers	Physical well-being
Company doctors	Physical well-being, psychological well-being

Human values

In this section, we describe in more detail which impact participants mentioned and the different human values that are implicated, based on the human values as described and identified by Friedman et al. (2013). For an overview of the impact mapped on stakeholders and affected human values, see Figure 1.

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). Regarding autonomy, in general, participants were concerned over their dependence on technology. If FRs become too dependent on the technology, their ability to operate will be substantially impaired when systems fail. In addition, a concern is staying in control of decisions.

Regarding dispatchers, three aspects were mentioned. AI might become more and more autonomous and take over several tasks of the dispatcher. This could (1) lead to less autonomy in their task execution, and (2) impact the decision-making power of the dispatcher, and (3) disproportional trust of the results or advise of AI systems, with dispatchers reflecting less on the situation and thus operating less autonomously.

Regarding fire fighters, it is expected that AI systems lead to new needs regarding capabilities and knowledge. This might (negatively) impact their ability to act and make decisions in their operations. Furthermore, AI systems will also have impact on their task decision-making power, depending on the level of autonomy of AI systems regarding decision-making. In addition, because possible access to historic data, AI systems can give advise on (long- and short-term) employability, which might impact fire fighters' control about own employability, with possibly others deciding on their performance and health. Participants are concerned that there will be less space to make decisions based on own insights, and that human aspects, context, and experience will be less taken into account when making decisions.

For incident commanders, participants mention that AI could lead to the feeling of not being in control of own decisions anymore. Participants were also worried regarding the influence of AI advise on the decisions of incident commanders. Also, they mentioned that the human aspects remain very relevant in decision-making; for example, optimal teams based on some objective measure are not always the best, as other goals of teams should be taken into account, such as that people can learn and grow.

No positive impact of AI systems on autonomy was mentioned.

Identity Identity refers to people's understanding of who they are over time, embracing both continuity and discontinuity over time (Friedman et al., 2013). Participants mentioned that AI systems will have impact on the capabilities, knowledge, and possibly level of education that is needed for the job (of incident commanders, fire fighters as well as dispatchers). For example, AI systems might become more and more autonomous and might take over decisions that at the moment are taken by dispatchers. The participants saw this as a threat for their understanding of their work. For fire fighters and incident commanders, participants also mentioned that a culture change is needed, as AI gives completely new possibilities, which means that operations will need to change (e.g., regarding gaining situation awareness first remotely by means of autonomous UGVs and UAVs, instead of through human operation). In addition, participants expect that new (commander) roles will be created, e.g., a role as specialist digital exploration.

	Autonomy	*Identity*	*Informed consent*	*Privacy*	*Physical well-being*	*Psychological well-being*	*Trust*
Incident commander	- Dependence on AI - Decision-making - Human aspects	- Culture change is needed			+ Can monitor safety of team + Better long-term employability + Safer incident response		
Fire fighter	- Ability to make decisions - Control on employability - Human aspects and experience - Dependence on AI	- Needed capabilities, knowledge, and level of education will change - Culture change is needed	- Unforeseen use of data	- Unforeseen use of data - Who has access to data? - Feeling of being closely monitored	+ Better insight in own health status + Less stress + Better health + Better long-term employability + Safer incident response - (Felt) unpredictability of UGVs - Physical impairment	+ Less stress - Less social environment	- Feeling of being closely monitored
Dispatcher	- Task execution - Decision making - Reflection	- Needed capabilities, knowledge, and level of education will change - Threat for work		- More access to sensitive data	+ Better insight in own health status + Safer incident response through faster alarming	+ Less PTSD + Less stress	
Citizen				- Personal data	+ More safety through better recognition of environmental effects	- Less personal contact	- Feeling of being monitored
Surroundings					+ More safety through better recognition of environmental effects		
Special ops					+ More safety through better insight in presence dangerous substances		
Incident researcher					+ More safety through better insight and analysis on prevention		
Company doctor					+ Better overview of health status + Better early detection	+ Better overview of stress and trauma	

Figure 1. Ethical matrix. Red cells denote negative impact on the corresponding value, green cells positive impact, and yellow cells positive as well as negative impact.

Informed consent Informed consent refers to "garnering people's agreement, encompassing criteria of disclosure and comprehension (for 'informed') and voluntariness, competence, and agreement (for 'consent')" (Friedman et al., 2013). The participants mentioned that the introduction of AI systems could have impact on informed consent. It might not be clear what exactly is being done with (personal) data that is collected of fire fighters and that, when available, this data could be used for purposes that were not foreseen. The participants mentioned that a possible solution is the introduction of personal data safes which are not generally accessible and that people can determine in which situations these data safes are accessible and to whom (e.g., in life-critical situations to their incident commanders).

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 privacy was mentioned for fire fighter, dispatchers as well as citizens. For fire fighters, impact on privacy is seen mostly regarding two aspects, i.e., that of proportionality and of surveillance. First, when collecting and analyzing (personal) data, it is not always clear what the goal is of doing so and what the data is actually used for. There must be clear agreements on what AI systems do with the data and who has access to it (and to the results of a possible analysis). In addition, participants mentioned that fire fighters in general resist having their data collected, as it gives them the feeling of being closely monitored.

Dispatchers will need to take privacy (even more) into account. They will have more access to, for example, cameras on roads during incidents, and can directly watch the situation to support their situation awareness. Also, they are in direct contact with other first response organizations, such as police, that have access (and share) information during incidents.

Regarding citizens, participants mentioned that with the introduction of AI systems, care needs to be taken with personal information of citizens, such as video pictures and gps data.

Physical well-being Physical well-being falls under the value *Welfare* as identified by Friedman et al. (2013); in this context, we understand it as physical health and safety. For all identified stakeholders, participants mentioned that AI systems could impact the value *physical well-being*. AI systems could have a positive impact on the physical safety and well-being of incident commanders, fire fighters, citizens, and the surroundings as it could lead to a better situation awareness because of additional relevant information (on the incident and on personnel). This information leads to a more effective task performance through faster, earlier, more stable, and more informed decision-making of all professional stakeholders. For fire fighters, dispatchers, and incident commanders, having more information on one's own situation and being able to monitor and analyse one's physical status (e.g., regarding stress) and

acting on the status could make the job healthier and lead to better long-term employability. This holds also for the company doctor, in addition to possible early detection of (work-related) illness. For special ops and incident researchers, AI systems might lead to better insight and analysis of dangerous effects or situations, which in turn might lead to a safer incident response for all stakeholders. Also, deploying drones increases the feeling of safety.

However, physical well-being could also be impacted negatively. For fire fighters, carrying sensors or AI systems might impair movement and mobility, leading to less safe performance. In addition, being with UGVs in the same environment can increase a feeling of unsafety.

Psychological well-being Psychological well-being also falls under the value *Welfare* as identified by Friedman et al. (2013). According to the participants, for dispatchers as well as for fire fighters, AI systems can lead to improved psychological well-being. The participants mentioned that stress and PTSD (for dispatchers) could occur less because of more effective operations and because of being able to monitor one's (stress) situation.

As possible negative impacts, participants mentioned that the fire service organisation is a very social environment. A concern is that this will change with the introduction of AI systems, as human aspects might become less important, which might lead to less (psychological) well-being of fire fighters. In addition, AI systems might lead to less contact between citizens and dispatchers (due to for example automatic alarming); however, personal contact is important for citizens, to be heard and supported in a stressful situation.

Trust Trust refers to expectations that exist between people who can experience good will, extend good will toward others, feel vulnerable, and experience betrayal (Friedman et al., 2013). The human value trust is mentioned by participants to be impacted by AI for two stakeholder groups, namely fire fighters and citizens, with respect to trust in the first response organization.

For fire fighters, to effectively use AI, more data will be collected, also from fire fighters. Participants mentioned that fire fighters generally distrust the collection of (personal) data and feel closely monitored. They worry that the data is used against them (e.g., proving that a performance is not up to par). In addition, no context is present in the data, which is important to explain, for example, deviation from protocol. For citizens, trust in first response can be decreased for the same reasons. For example, the position of emergency callers is available for dispatchers, as well as of other people at the incident location. Monitoring and following these citizens (through AI systems) might enable a more effective operation; however, it could decrease trust in first response organisations due to a feeling of being monitored.

AI-related values

The participants of the focus group sessions mentioned several other values regarding the introduction of AI for fire services that are not part of the human values that are often implicated by ICT systems as identified and described by Friedman et al. (2013). This is not surprising, as Friedman's list of values relates to ICT systems in general and not AI systems in particular. The other values mentioned are *Accountability*, *Reliability*, *Trust in AI systems*, *Security*, *Transparency*, and *Appropriate training*. These values were mentioned in the context of what AI systems should adhere to, or how AI systems should function and were mostly mentioned independent of stakeholders, but as general aspects that should be taken into account during the design, development, and deployment of AI.

Accountability Accountability is seen in the context of liability (legal responsibility) as well as social responsibility. Participants are concerned that the legal responsibility is not clearly specified; aspects as liability when (wrong) decisions are taken (by the AI system as well as by the FR when being advised differently by an AI system) need to be taken into account when introducing AI systems. Also, for FRs, it is very important to be accountable for the decisions taken during operations, e.g., to be able to explain decisions after the operation. This also needs to be the case when AI systems are integrated into the operation.

Reliability Participants mentioned that it is important that AI systems are reliable, both in the results they present (e.g., in all situations, they should present working results) as well as in availability of the systems (e.g., they should not fail in high-stress or dangerous situations). This corresponds to the key requirement *Technical robustness and safety* of the HLEG (AI HLEG, 2018).

Trust in AI systems Participants mentioned that (appropriate) trust in AI systems is important. Topics that were mentioned regarding trust in AI systems were (1) understanding of system as the basis for trust, (2) *appropriate* trust, (3) dependence on systems. Regarding understanding of the system, participants mentioned that it is important to understand what the system is doing (e.g., predicting the next actions of a UGV), how the AI got to a result or advise, and how certain the advise is. Regarding appropriate trust, 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 on the advise of the AI) or of too little trust (and not taking the advise of the AI sufficiently into account). Participants were also concerned that if the AI system has some kind of autonomy (e.g., when recognizing incidents and sending relevant information about the incident when alarming first response), the FRs have no choice other than to trust the system, and thus depend on the system, as there is no time to validate all information before acting.

The participants also stressed that an incremental introduction of AI is necessary to learn about the capabilities of the AI system and how to integrate it into work processes and build up trust. This also is important because often, AI systems in practise do not work as well as in theory, which leads to distrust and less support.

Security Regarding security, participants mentioned that it is very important that data is kept at a secure place and that access control is in place. Misuse of data needs to be avoided; cybercrime is a danger.

Transparency Regarding transparency, participants mentioned several important aspects. First of all, it is important to have an understanding of the results of the AI system (e.g., advice) and the reasons why the AI system comes to these results. In addition, it is important to have an idea on the certainty of the results. Also, explanations of the advise are important, ideally taking the context into account (e.g., which context factors improve the certainty). Regarding UGVs, participants mentioned that it is important to have insight into the plan/goal of the robot, to be able to predict behaviour.

Appropriate training Regarding appropriate training, participants mentioned two topics: (1) it is of utmost importance that FRs learn how to work with AI systems, as most do not currently have sufficient IT knowledge, and (2) that it is important that FRs explicitly keep training on how to work effectively without AI systems, so that they are able to operate if the system fails and are able to evaluate advise from AI systems regarding applicability.

IMPLICATIONS ON DESIGN OF AI SYSTEMS FOR FR

In this section, we describe the implications of the results as described in the previous section on the design of responsible AI for fire services. As shown in Figure 1, positive as well as negative impact has been identified during the focus group sessions for the different human values. According to the participants, several human values are impacted (mostly) positively for the different stakeholders (although there is no human value that is only affected positively), i.e., *physical well-being* and *psychological well-being*. Several human values are impacted only negatively for the different stakeholders, i.e., *autonomy*, *identity*, *informed consent*, *privacy*, and *trust*. Particular attention should go to these values in the design-, development-, and deployment process of AI systems; 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. To do this, the AI-related values that have been identified can help; these were values that the participants said that AI systems should adhere to. For example, having design requirements regarding the values *accountability*, *transparency*, and *appropriate training* could help to reduce the expected negative impact on *autonomy*. Regarding *accountability*, understanding the responsibility distribution between an AI system and human would help a FR's ability to act and make decisions; regarding *transparency*, understanding advise of an AI system in the current context and being able to interpret its reliability lessens dependence on the results of the AI system; the same holds for being *trained* in how to work with and without the AI system. The translation from human values to design requirements is not straightforward (Aldewereld & Mioch, 2021). However, the values form a good basis to take into account in the design and development process of specific AI applications. For each specific AI application, the values should be re-evaluated for applicability and, together with stakeholders, value-based design requirements should be specified that promote positive impact and limit negative impact on these values. For our current research, design requirements can only be very abstract as we have not identified a specific AI system that is the subject of this research, but AI systems in general. As a proof of concept, we give an example of a translation of the value *Autonomy* into possible general design requirements, see Table 3.

Table 3. Examples of design requirements for a selection of the identified impact of AI systems on the value *Autonomy*.

Impact of AI system on Autonomy	*Design Requirements*
Dependence on AI	Personnel will[1] be extensively trained on how the AI system works and how to interact with the system during operations Personnel will be extensively trained on how to work. effectively without the AI system in case of failure and to be able to evaluate advise from the AI system. The AI system will be able to explain why a particular result or advise is given. This explanation will take the context into account and be adapted towards the particular task and role of the FR.
Decision-making	The human will stay in control of (operational) decisions. The role of the AI system in the task execution will be clear and integrated into procedures. Accountability and responsibility will be explicitly discussed and set for each AI system that is introduced.
Human aspects and experience	Protocols will be specified carefully regarding decision-making with AI systems, to make sure that human aspects and experience can be taken into account. The AI system will be able to learn from human experts.

[1] We use 'will' in the design requirements specification instead of 'shall' or 'should', as the latter imply a normative load.

CONCLUSION AND DISCUSSION

In this paper, we have made a first step towards a generic framework of ethical aspects of AI systems for FR. We held three focus group sessions with different stakeholders, discussing several value-sensitive scenarios and the expected impact of the described AI systems on different stakeholders. Based on the results of the sessions, we identified relevant stakeholders and values that could be impacted by the introduction of AI systems. Mapping the expected impact onto the ethical matrix gave a visual help to show clearly which values are supported by AI technology and which might be negatively affected. We have also given an example of how these values van be translated into design requirements.

The number of participants in the focus group sessions was relatively limited. Nonetheless, the participants were diverse in the function they hold in the fire organisation, which leads us to believe that the analyses give good insights into relevant values and expected impact of AI applications for fire services.

During the first focus group session we realized that to be able to identify possible impact of AI systems, a general knowledge of AI systems is needed. For that reason, we added an introduction into AI systems to the subsequent sessions. In addition, we learned that example scenarios support participants in their identification of impact. During the subsequent focus group sessions, we used scenarios to identify stakeholders and human values for fire services. The scenarios were well grounded as they were based on input from domain experts from the first focus group session, AI experts (prediction of possible technology), and literature (relevant expected values). We introduced two scenarios with different AI applications and different direct stakeholders to broaden results. The goal of the scenarios was to support the participants in thinking about AI technology and its possibilities, without limiting it to a specific implementation. The choice of these specific scenarios led to a first inventory of relevant human values, which we expect for future scenarios will be refined, extended, and if necessary amended. The chosen scenarios led mostly to the identification of impact on stakeholders within the fire services; this was intentional. In future work, 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, as we expect that additional impact will be identified, leading to different affected values.

In the FR domain, previous research has identified relevant stakeholders and core ethically relevant themes and values for the application of rescue robots. For example, in Harbers et al.'s selection of most relevant stakeholders included victims, local authorities, electrical company, press, and observers (next to fire fighters, police, and ambulance) (Harbers et al., 2017). These stakeholders were also identified during our focus group sessions, but not prioritized for AI applications and our scenarios. For rescue robots, Harbers et al. (2017) identified the values personal safety, safety of others, access to information, well-being, effectiveness, ease of use, authority for fire fighters. Although these values do not directly correspond to the values that resulted from our sessions as we based our value definition on Friedman et al. (2013), most of them relate to them. Ease of use has not been

mentioned during our sessions. Battistuzzi et al. (2021) identified core ethically relevant themes by means of a scoping review, namely fairness and discrimination, false or excessive expectations, labor replacement, privacy, responsibility, safety, and trust. Most of these topics were mentioned during our focus group sessions; however, fairness and discrimination and labor replacement were not mentioned. During our focus groups, impact on the values *autonomy* and *identity* were mentioned clearly and were seen as very important. These values have not been identified by (Harbers et al., 2017) and (Battistuzzi et al., 2021). These values might be (more) affected by AI systems for decision support than by rescue robots, as building situation awareness and decision making are core tasks for FRs. Several human values that were identified in this research were impacted only negatively for the different stakeholders; in future research, it should be further investigated why this was the case.

In our research, some values that were mentioned by the participants correspond (partly) to some of the key requirements as promoted, for example, by the EU HLEG for AI (e.g., transparency) (AI HLEG, 2018). When comparing the values that were mentioned by the participants at the focus group sessions with the EU HLEG ethical principles and key requirements, it becomes clear that several requirements particularly relevant for AI systems were not mentioned by the participants, e.g., diversity, non-discrimination, and fairness, or environmental well-being.

There is a general consensus that AI technology raises ethical issues that are not raised by more conventional ICT technology (Floridi et al., 2018). For that reason, the original VSD list of values does not suffice for AI (Umbrello & Van de Poel, 2021) but should be supplemented by ethical principles that ensure that typical AI ethical issues are addressed. Additional to individual values, organizational and societal values and corresponding norms need to be taken into account to build socially responsible AI systems. The results of our bottom-up, empirical approach of identifying context-specific values should be supplemented by two other normative sources of values i.e., (1) values promoted by the design, such as by deriving from the UN's Sustainable Development Goals (Guterres, 2020) or from AI for Social Good principles; (2) values respected by the design, particularly values identified in relation to AI (e.g., ethical principles identified by the EU HLEG for AI (AI HLEG, 2018)).

We foresee that AI technology will not only impact the stakeholders that the participants identified as most relevant during the focus group sessions, but that AI will have impact on the first response organizations in their totality, changing roles and responsibilities, but also the way resources are allocated, how training will take place, and what is considered expertise (as is argued also for other domains, such as health care (Van Wynsberghe & Li, 2019)). Some of the impact identified by the participants supports this (e.g., expected impact on training needs, needed capabilities for first response work, and the need for a culture change). This makes it even more important to introduce AI systems responsibly, decreasing negative impact of the introduction of AI systems explicitly during the design and development process, evaluating AI systems not only in the direct interaction with the direct stakeholders, but also evaluating effects on the organizations.

The values identified in this research function as a basis for a general value set and will be iteratively built up during future research, together with a related requirements list, as well as the (related) specific values and (instantiated) requirements for critical scenarios. We started applying this research to a particular AI application for fire services (an AI-based decision-support system) (Mioch et al., 2024), to further specify values for this particular AI application, map these values onto design requirements, and operationalize them for the AI application to investigate practicability and applicability of the general framework. In addition, we will also continue with conceptual investigations by integrating values found from normative sources (such as AI for Social Good) into the results of this research, in addition to doing a scoping literature review on ethical aspects of AI systems applied to emergency response and integrating these findings into the ethical framework.

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. (2018). Ethics Guidelines for Trustworthy AI. *European Commission*, 14–16.

Akata, Z., Balliet, D., de Rijke, M., Dignum, F., Dignum, V., Eiben, G., Fokkens, A., Grossi, D., Hindriks, K., Hoos, H., Hung, H., Jonker, C., Monz, C., Neerincx, M., Oliehoek, F., Prakken, H., Schlobach, S., van der Gaag, L., van Harmelen, F., ... Welling, M. (2020). A Research Agenda for Hybrid Intelligence: Augmenting Human Intellect With Collaborative, Adaptive, Responsible, and Explainable Artificial Intelligence. *Computer*, *53*(8), 18–28. https://doi.org/10.1109/MC.2020.2996587

Aldewereld, H., & Mioch, T. (2021). Values in design methodologies for AI. *International Conference on Advanced Information Systems Engineering*, 139–150.

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

Blasimme, A., & Vayena, E. (2020, July). The Ethics of AI in Biomedical Research, Patient Care, and Public Health. In M. D. Dubber, F. Pasquale, & S. Das (Eds.), *The Oxford Handbook of Ethics of AI*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190067397.013.45

Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid Intelligence. *Business & Information Systems Engineering*, *61*(5), 637–643. https://doi.org/10.1007/s12599-019-00595-2

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

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

Galliott, J., & Scholz, J. (2020, July). The Case for Ethical AI in the Military. In M. D. Dubber, F. Pasquale, & S. Das (Eds.), *The Oxford Handbook of Ethics of AI*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190067397.013.43

Guterres, A. (2020). The sustainable development goals report 2020. *United Nations publication issued by the Department of Economic and Social Affairs*, 1–64.

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
cited By 9.

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

IFAFRI. (2019, September). Capability Gap 9 "Deep Dive" Analysis.

Mepham, B., Kaiser, M., Thorstensen, E., Tomkins, S., & Millar, K. (2003). *Ethical Matrix Manual* (tech. rep.). LEI, Wageningen UR.

Mioch, T., Aldewereld, H., & Neerincx, M. A. (2024). Human values for responsible decision-support for fire services. In B. Penkert, B. Hellingrath, A. Widera, H. Speth, M. Middelhoff, K. Boersma, & M. Kalthöner (Eds.), *Proceedings of the 21st ISCRAM conference*.

Mioch, T., Sterkenburg, R., Beuker, T., & Neerincx, M. A. (2021). Actionable Situation Awareness: Supporting Team Decisions in Hazardous Situations. *Proceedings of the 18th International Conference on Information Systems for Crisis Response and Management, ISCRAM 2021, Blacksburg 23 May 2021 through 26 May 2021*, 62.

Morley, J., Machado, C. C. V., Burr, C., Cowls, J., Joshi, I., Taddeo, M., & Floridi, L. (2020). The ethics of AI in health care: A mapping review. *Social Science & Medicine*, *260*, 113172. https://doi.org/10.1016/j.socscimed.2020.113172

Murphy, K., Di Ruggiero, E., Upshur, R., Willison, D. J., Malhotra, N., Cai, J. C., Malhotra, N., Lui, V., & Gibson, J. (2021). Artificial intelligence for good health: A scoping review of the ethics literature. *BMC Medical Ethics*, *22*(1), 14. https://doi.org/10.1186/s12910-021-00577-8

Radianti, J., Dokas, I., Boersma, K., Noori, N. S., Belbachir, N., & Stieglitz, S. (2019). Enhancing Disaster Response for Hazardous Materials Using Emerging Technologies: The Role of AI and a Research Agenda. In J. Macintyre, L. Iliadis, I. Maglogiannis, & C. Jayne (Eds.), *Engineering Applications of Neural Networks*

(pp. 368–376, Vol. 1000). Springer International Publishing. https://doi.org/10.1007/978-3-030-20257-6_31

Schwartz, S. H. (2012). An Overview of the Schwartz Theory of Basic Values. *Online Readings in Psychology and Culture*, *2*(1). https://doi.org/10.9707/2307-0919.1116

Seeber, I., Bittner, E., Briggs, R. O., de Vreede, T., de Vreede, G. J., Elkins, A., Maier, R., Merz, A. B., Oeste-Reiß, S., Randrup, N., Schwabe, G., & Söllner, M. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information and Management*, *57*(2), 103174. https://doi.org/10.1016/j.im.2019.103174 Publisher: Elsevier.

Umbrello, S., & Van de Poel, I. (2021). Mapping value sensitive design onto ai for social good principles. *AI and Ethics*, *1*(3), 283–296.

van de Poel, I. (2015). Design for Values in Engineering. In J. van den Hoven, P. E. Vermaas, & I. van de Poel (Eds.), *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains* (pp. 667–690). Springer Netherlands. https://doi.org/10.1007/978-94-007-6970-0_25

van der Stappen, E., & Steenbergen, M. v. (2020). The Ethical Matrix in Digital Innovation Projects in Higher Education. *BLED 2020 Proceedings*.

Van Wynsberghe, A., & Li, S. (2019). A paradigm shift for robot ethics: From hri to human–robot–system interaction (hrsi). *Medicolegal and Bioethics*, 11–21.

Wasilow, S., & Thorpe, J. B. (2019). Artificial Intelligence, Robotics, Ethics, and the Military: A Canadian Perspective. *AI Magazine*, *40*(1), 37–48. https://doi.org/10.1609/aimag.v40i1.2848

Wright, D., Finn, R., Gellert, R., Gutwirth, S., Schütz, P., Friedewald, M., Venier, S., & Mordini, E. (2014). Ethical dilemma scenarios and emerging technologies. *Technological Forecasting and Social Change*, *87*, 325–336. https://doi.org/10.1016/j.techfore.2013.12.008