Supporting Responsible Human-AI Cooperation for Emergency Responders

Mioch, Tina

DOI
10.3233/FAIA250690

Publication date
2025

Document Version
Final published version

Published in
HHAI 2025 - Proceedings of the 4th International Conference on Hybrid Human-Artificial Intelligence

Citation (APA)
Mioch, T. (2025). Supporting Responsible Human-AI Cooperation for Emergency Responders. In D. Pedreschi, M. Milano, I. Tiddi, S. Russell, C. Boldrini, L. Pappalardo, A. Passerini, & S. Wang (Eds.), *HHAI 2025 - Proceedings of the 4th International Conference on Hybrid Human-Artificial Intelligence* (pp. 552-559). (Frontiers in Artificial Intelligence and Applications; Vol. 408). IOS Press. https://doi.org/10.3233/FAIA250690

HHAI 2025
D. Pedreschi et al. (Eds.)

doi:10.3233/FAIA250690

Supporting Responsible Human-AI Cooperation for Emergency Responders

Tina MIOCH [a,b,1]
[a] *University of Applied Sciences Utrecht*
[b] *Delft University of Technology*
ORCiD ID: Tina Mioch https://orcid.org/0000-0003-2457-5872

Abstract. Emergency responders (ERs) work in dynamic, complex, and unpredictable environments. The integration of increasingly advanced sensor and artificial intelligence technology affects the work processes and decision-making of ERs and leads to new challenges regarding efficient and responsible human-AI cooperation. For responsible human-AI cooperation, AI systems should support relevant human values; however, it is not clear which values these are and how they can be supported. This paper outlines an investigation into developing a framework for ER-AI cooperation to advance the building of shared situation awareness and decision-making, while taking the underlying values into account, and presents first results of relevant values and design requirements.

Keywords. Responsible Human-AI cooperation, Emergency response, Decision-support, Value-sensitive design

1. Introduction

Emergency response organizations work in environments and situations that are complex, dynamic, and unpredictable. Obtaining actionable situation awareness (SA) to make fast and effective decisions is of utmost importance. Actionable SA includes observing the environment, assessing situations, predicting plausible future scenarios, and being able to determine which of all possible actions should be executed [1]. To support gaining SA in dynamic and complex situations, more and more AI systems are being introduced in emergency response organizations, e.g., for fire detection [2] and real-time surveillance of the incident scene [3]. 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, advice on decisions based on previous incidents or monitoring emergency responders (ERs) safety by tracking a ERs heartbeat to look for signs of overexertion and issuing warnings based on this information. However, the introduction of AI technology leads to the challenge of integrating these AI systems in the decision-making processes in a responsible way by properly addressing (public) values and public function. 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

[1] Corresponding Author: Tina Mioch, tina.mioch@hu.nl

developing hybrid human-AI systems in which AI technology and humans cooperate in a way that creates synergy [4,5]. In such hybrid human-AI systems, complex goals can be achieved by combining humans and AI technology that collectively work on shared objectives with complementary capabilities that, when combined, augment each other [6]. The development and application of these hybrid human-AI systems need to be done responsibly [7], e.g., regarding possible biases in (training) data sets and privacy aspects (e.g., (personal) data on ERs 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 [8,9] and the military domain [10]. However, in the field of emergency response, it seems that ethics with regard to the application of AI has not yet been much addressed.

In this research project, we work on a generic framework to support responsible ER-AI cooperation, including insights into relevant ethical aspects and design requirements to take into account when developing responsible AI systems for emergency response, design patterns to describe ER-AI cooperation, and measurements and corresponding tools and methods for evaluating responsible human-AI cooperation.

2. Background

AI is a system technology that will fundamentally change society and it is not yet clear how AI can be implemented in a responsible and ethical way [11]. In recent years, several guidelines have been developed for the responsible application of AI [7,12]. However, these guidelines do not provide practitioners with sufficient tools for the design and deployment of responsible AI [13]. This leaves emergency response organizations with the challenge of integrating these AI decision-support systems responsibly into decision-making processes. In addition, emergency response organizations foresee a change in tasks, roles, and responsibilities through the deployment of (intelligent) decision-support systems, with ERs passing (some) responsibility and tasks to AI systems. It is not clear how these changes will be implemented in current work processes and decision responsibility and hierarchy, and how human-AI cooperation should take shape.

Johnson and Vera [14] developed a framework for organizing many of the important concepts associated with human-AI cooperation, focusing on interdependence between the different actors. It provides a general structure for use in the development of AI systems that cooperate well with humans. The framework also shows the limitations of current cooperative intelligence of AI systems. For example, current AI systems often collect large amounts of data about a situation but have limited self-awareness of its own limitations and awareness of the other's capabilities and little strategic behavior such as integration into human workflow. From an interaction perspective, Team Design Patterns (TDPs) have been presented to provide the means to capture human-AI teaming processes [15]. A design pattern is an evaluated and abstracted solution to a common problem [16]. These TDPs describe how humans and AI systems cooperate with each other, in a task-independent way that includes requirements, advantages, and disadvantages [17,18]. However, TDPs have not yet been applied and evaluated for real and dynamic world contexts with large incomplete and uncertain information [19].

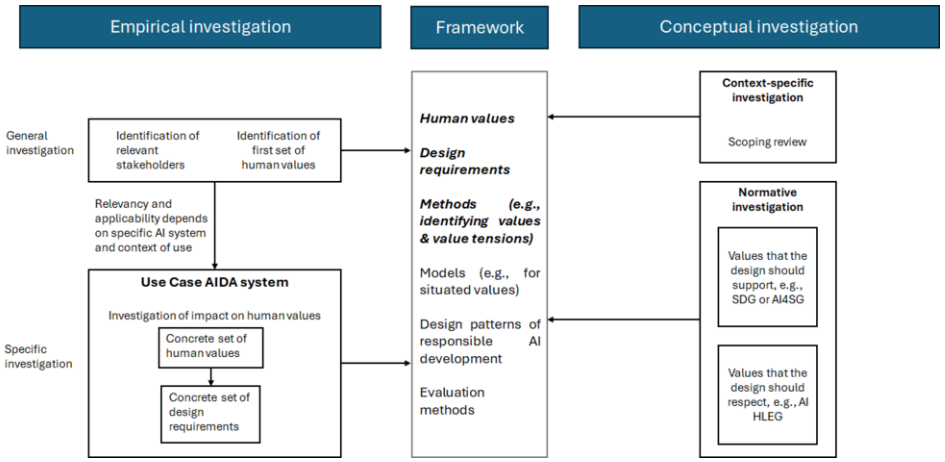

Figure 1. Overview of activities in study 1 and relation to general framework.

3. Research question

This leads to the following research question: *How can a collaborative decision-support framework for ER-AI teams advance the building of shared situation awareness and team performance during incident control in complex and dynamic situations, while taking underlying values into account?*

Investigating this research question comprises the following steps:

1. Identify the values and context factors that should be taken into account in the design, implementation, and deployment of AI decision-support systems for ER teams in complex and dynamic incidents.
2. Investigate how cooperation can be supported between ERs and AI systems regarding their roles, responsibilities, and mutual learning while taking into account the underlying values.
3. Identify methods that contribute to the evaluation of AI-human cooperation during complex and dynamic incidents.

4. Method and approach

This section describes, for each step identified in Section 3, the research approaches and activities that we will take to investigate the topic. We are inspired by the Value Sensitive Design (VSD) methodology, which accounts for human values throughout the design process [20]. 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 investigate values in technology, such as direct and indirect stakeholder analysis and value scenarios, and is thus particularly useful for our research.

4.1. Study 1:Responsible and value-sensitive requirements of AI decision-support systems for ERs

To investigate responsible and value-sensitive requirements of AI systems for ERs, we first conceptually and empirically investigate the values and other ethical aspects that are affected by AI systems in emergency response, followed by an analysis of the values to identify value-sensitive requirements for human-AI cooperation. For the conceptual investigation, we perform a scoping literature review on ethical issues that may emerge when AI systems are applied to emergency response and integrate the results with other sources of ethical norms, such as, e.g., EU guidelines. For the empirical investigation, first, relevant stakeholders are explored and relevant values are identified that are impacted by AI in general, followed by an investigation into a specific AI application. Based on these results, a requirement analysis and specification is performed. These activities lead to the formulation of a framework of ethical aspects for the introduction of AI systems in emergency response, which will provide insight into the relevant ethical aspects to take into account when developing AI systems for ERs. For an overview of the activities of study 1 and the relation to the general framework, see Figure 1.

4.2. Study 2: Designing Responsible Cooperation between the ER and AI system

To design responsible cooperation between ERs and AI systems, we investigate how ER-AI cooperation can be supported. The ethical framework for AI systems in emergency response (result of study 1) is used to determine the focus of the design activities in this study. Building on existing work of TDPs [21], a first design of a framework for supporting ER-AI cooperation is made, e.g., of mutual awareness of responsibilities and capabilities in complex and dynamic environments. This first design will be a low-fidelity paper-based prototype and will be evaluated during expert group sessions with stakeholders from emergency response organizations and scientific expert reviews. Requirements and design criteria are improved based on the results, and a more 'experienceable' model (e.g., a value-oriented mock-up) is designed (second design cycle). This mock-up is evaluated through an expert review on the validity of the design patterns and with ERs through usability tests.

4.3. Study 3: Evaluation of responsible human-AI cooperation

We iteratively develop and adapt objective and subjective measures and the corresponding tools and methods to evaluate responsible human-AI cooperation for this domain during the design cycles of study 2. Semi-structured interviews are conducted with different stakeholders on the application of the methods (including the evaluation criteria, measurements, and steps in the evaluation process), and observations are made during the evaluations. Based on these results, methods and tools are refined. This results in applied methods and tools for the evaluation of responsible ER-AI cooperation in complex and dynamic situations.

5. First results

Currently, we are finishing study 1. Regarding the conceptual investigation, we are working on the scoping literature review on ethical issues that may emerge when AI systems

Table 1. Examples of identified impact, the impacted values, and the corresponding module, as reported in Mioch et al. [24].

Value	Impact	Module
Autonomy	Overreliance on AI	Video analysis module, Material prediction module
Well-being	Fewer accidents when driving	Material prediction module
	More PTSD because of availability of incident videos	Video analysis module

are applied to emergency response. We are currently analyzing the results of this literature review; preliminary results suggest that most of the literature on ethical aspects of AI systems in emergency response is related to rescue robotics, with a very limited set of literature working on ethical aspects of other AI systems (such as decision support). The empirical investigation consisted of two steps; first, an investigation of values that are impacted by AI systems in general, identifying relevant stakeholders and a first set of human values that are impacted by AI systems for these stakeholders; second, an investigation of relevant human values and other ethical aspects 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 concentrated on two modules, the *video analysis module* that analyzes drone video streams and detects humans [22] and a *material prediction module* that, based on correlations found in past incident data, predicts the fire services material needed for an incident.

In the first iteration on impacted human values by AI systems in general, we conducted several focus group sessions with fire service personnel to identify key human values. Some values were found to be (mostly) positively affected for the different stakeholders (although there was no value that was only positively affected), i.e., *physical well-being* and *psychological well-being*. Several values were found to be only negatively affected for the different stakeholders, i.e., *autonomy*, *identity*, *informed consent*, *privacy*, and *trust*. Other values that were identified to be affected by AI systems were *accountability*, *transparency*, *reliability*, *security*, and *appropriate training* [23].

In the second iteration on affected human values by the AIDA system, we held 9 expert group sessions with a total of 36 fire service employees, of which 15 (head) incident commanders, 17 dispatchers, 3 fire fighters, and 1 emergency coordinator, and discussed the impact of AIDA on different stakeholders. The identified impact of AI systems was analyzed through thematic analysis and mapped onto affected underlying values. Preliminary results show that (1) previously identified values (for general AI systems) are implicated by either of the two AI modules, e.g., *Identity*; (2) impact on values such as *Well-being* and *Autonomy* has been mentioned for both modules, and (3) there are differences between the two modules on how and which values are impacted [24]. In Table 1, examples of impact and corresponding value identified for the two modules are presented.

Currently, we are also integrating the results of the different investigations into a general framework, identifying design requirements to lessen the negative impact and strengthen the positive impact of AI systems (see for examples of design requirements Table 2). These requirements should be considered for the design, development, and deployment of responsible collaborative decision-support of ER teams in complex and dynamic environments. We will also report on lessons learned of methods to identify impacted values and value tensions.

Table 2. Examples of design requirements for one identified impact of AI systems on the value *Autonomy* [23].

Impact	*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 ER.

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

6. Future work

As next steps, we will start designing human-AI cooperation that supports identified human values (e.g., autonomy), based on design requirements that have been identified in the empirical investigation. We will choose a specific value to focus on and analyze how this value is supported in the current setup of the decision-support system AIDA. We will then develop TDPs to improve the cooperation between ER and AIDA to better support the selected value. This will be done iteratively, designing, evaluating, and improving the cooperation in several cycles, also developing methods to evaluate whether the particular value has been supported. We believe that our results are also highly relevant to other domains that involve decision-making under time pressure and within dynamic, complex environments with incomplete information, such as military, police, and healthcare settings, and we will apply the results to a use case in another domain.

This research will result in a framework for collaborative decision-support to advance the building of shared SA and team performance, while taking underlying values into account, and applied methods and tools for the evaluation of ER-AI cooperation in complex and dynamic environments. In addition, this research will contribute to the desire expressed by emergency response organizations for more insight into how cooperation between AI systems and ERs can be designed to optimally support decision-making, taking relevant values into account. By using an iterative design methodology and evaluation of a prototype in complex and realistic situations with stakeholders, AI development for practice will be accelerated; awareness within the organization will be created regarding advantages and challenges of AI decision-support systems and of the necessity of a responsible and context-dependent implementation of AI. Furthermore, the developed prototype will be used for further development of responsible AI systems for ERs.

Acknowledgements

I would like to thank my promotor Mark A. Neerincx, co-promotor Luciano Cavalcante Siebert, daily supervisor Huib Aldewereld, and advisor Stefan Leijnen for their support and advice. I also would like to thank the Safety Regions *Rotterdam-Rijnmond* and *Zuid-Holland-Zuid* and the *Gezamenlijke Brandweer* for their participation in this research. This research is supported by the University of Applied Sciences Utrecht (HU) through a '*promotievoucher*'.

References

[1] Endsley MR. Toward a theory of situation awareness in dynamic systems. Human Factors. 1995;37(1):32-64.

[2] Talaat FM, ZainEldin H. An improved fire detection approach based on YOLO-v8 for smart cities. Neural Computing and Applications. 2023 Oct;35(28):20939-54. Available from: https://doi.org/10.1007/s00521-023-08809-1.

[3] Martinez-Alpiste I, Golcarenarenji G, Wang Q, Alcaraz-Calero JM. Search and rescue operation using UAVs: A case study. Expert Systems with Applications. 2021 Sep;178:114937. Available from: https://www.sciencedirect.com/science/article/pii/S095741742100378X.

[4] Seeber I, Bittner E, Briggs RO, de Vreede T, de Vreede GJ, Elkins A, et al. Machines as teammates: A research agenda on AI in team collaboration. Information and Management. 2020;57(2):103174. Available from: https://doi.org/10.1016/j.im.2019.103174.

[5] Akata Z, Balliet D, de Rijke M, Dignum F, Dignum V, Eiben G, et al. A Research Agenda for Hybrid Intelligence: Augmenting Human Intellect With Collaborative, Adaptive, Responsible, and Explainable Artificial Intelligence. Computer. 2020 Aug;53(8):18-28. Available from: https://ieeexplore.ieee.org/document/9153877/.

[6] Dellermann D, Ebel P, Söllner M, Leimeister JM. Hybrid Intelligence. Business & Information Systems Engineering. 2019 Oct;61(5):637-43. Available from: http://link.springer.com/10.1007/s12599-019-00595-2.

[7] High Level Independent Group on Artificial Intelligence (AI HLEG). Ethics Guidelines for Trustworthy AI. Euorpean Commission. 2018:14-6. Available from: https://ec.europa.eu/newsroom/dae/document.cfm?doc{_}id=60419.

[8] Murphy K, Di Ruggiero E, Upshur R, Willison DJ, Malhotra N, Cai JC, et al. Artificial intelligence for good health: a scoping review of the ethics literature. BMC Medical Ethics. 2021 Feb;22(1):14. Available from: https://doi.org/10.1186/s12910-021-00577-8.

[9] Morley J, Machado CCV, Burr C, Cowls J, Joshi I, Taddeo M, et al. The ethics of AI in health care: A mapping review. Social Science & Medicine. 2020 Sep;260:113172. Available from: https://www.sciencedirect.com/science/article/pii/S0277953620303919.

[10] Galliott J, Scholz J. The Case for Ethical AI in the Military. In: Dubber MD, Pasquale F, Das S, editors. The Oxford Handbook of Ethics of AI. Oxford University Press; 2020. p. 0. Available from: https://doi.org/10.1093/oxfordhb/9780190067397.013.43.

[11] WRR. Opgave AI. De nieuwe systeemtechnologie - Rapport - WRR; 2021. Type: rapport. Available from: https://www.wrr.nl/publicaties/rapporten/2021/11/11/opgave-ai-de-nieuwe-systeemtechnologie.

[12] ECP. Artificial Intelligence Impact Assessment (English version). ECP; 2019. Available from: https://ecp.nl/publicatie/artificial-intelligence-impact-assessment-english-version/.

[13] Schiff D, Rakova B, Ayesh A, Fanti A, Lennon M. Explaining the Principles to Practices Gap in AI. IEEE Technology and Society Magazine. 2021 Jun;40(2):81-94.

[14] Johnson M, Vera AH. No Ai is an island: The case for teaming intelligence. AI Magazine. 2019;40(1):16-28.

[15] Van Diggelen J, Johnson M. Team design patterns. HAI 2019 - Proceedings of the 7th International Conference on Human-Agent Interaction. 2019:118-26.

[16] Alexander C. A pattern language: towns, buildings, construction. Oxford university press; 1977.

[17] Van Diggelen J, Neerincx M, Peeters M, Schraagen JM. Developing effective and resilient human-agent teamwork using team design patterns. IEEE intelligent systems. 2018;34(2):15-24.

[18] Van Zoelen E, Mioch T, Tajaddini M, Fleiner C, Tsaneva S, Camin P, et al. Developing team design patterns for hybrid intelligence systems. In: HHAI 2023: Augmenting Human Intellect. IOS Press; 2023. p. 3-16. Available from: https://ebooks.iospress.nl/volumearticle/63319.

[19] Harbers M, Neerincx MA. Value sensitive design of a virtual assistant for workload harmonization in teams. Cognition, Technology & Work. 2017;19(2-3):329-43.

[20] Friedman B, Hendry DG. Value sensitive design: Shaping technology with moral imagination. Mit Press; 2019.

[21] van der Waa J, van Diggelen J, Siebert LC, Neerincx M, Jonker C. Allocation of Moral Decision-Making in Human-Agent Teams: A Pattern Approach. In: International Conference on Human-Computer Interaction; 2020. p. 203-20.

[22] Heemskerk J, Mioch T, Maathuis H, Aldewereld H. Long-Range Human Detection in Drone Camera Images. Proceedings of the International ISCRAM Conference. 2024 May. Available from: https://ojs.iscram.org/index.php/Proceedings/article/view/75.

[23] Mioch T, Aldewereld H, Neerincx MA. Empirical Investigation of Values Affected by AI Systems for Fire Services. Proceedings of the International ISCRAM Conference. 2024 May. Available from: https://ojs.iscram.org/index.php/Proceedings/article/view/17.

[24] Mioch T, Aldewereld H, Neerincx MA. Human Values for Responsible Decision-Support for Fire Services. Proceedings of the International ISCRAM Conference. 2024 May. Available from: https://ojs.iscram.org/index.php/Proceedings/article/view/79.