UNIVERSITY OF TWENTE.

Faculty of Behavioural, Management and Social Sciences

Designing Human-Agent-Teaming for First Responders

Tatjana Beuker

Master Thesis

November 2021

Supervisors:
Prof. Dr. Frank van der Velde
Dr. Simone Borsci
MSc. Tina Mioch

In cooperation with:

TNO innovation for life

assistance

Abstract

To assist First Responders (FRs) during the mitigation of disasters that involve chemical hazardous substances, a technological advanced system, the Chemical Hazard Tool, is being developed. It will display information about a current and predicted gas cloud distribution by using constantly updated input from meteorological services, chemical sensors, and FRs. In this thesis we assess how the teamwork between the FR and the Chemical Hazard Tool, representing a Human-Agent Team (HAT), should be designed to facilitate decision making during the mitigation of dynamic mission evolvements. For this purpose, we created three Team Design Patterns (TDPs), each assigning different roles and responsibilities to the FR and the Chemical Hazard Tool. The three TDPs are named after the role the Chemical Hazard Tool takes on in each of the collaboration styles: *Informing Agent*, *Advising Agent* and *Deciding Agent.* The three TDPs were evaluated by FRs (N = 19) in an online survey which showed a low fidelity simulation of how the collaboration with each AI agent would look like. Results of this formative evaluation show that all three TDPs have their legitimacy and that there is no uniform consensus on the most suitable option. Preferences for TDPs indicated by the FRs varied depending on the task at hand, the circumstances of the disaster and the specialization of the FR. This leads to recommending a design solution in which the Chemical Hazard Tool can change from one TDP to the other depending on the FR and the decision the Chemical Hazard Tool is assisting with. Additionally, FRs indicated that trust, reliability and sufficient explanation of the underlying decision model are important factors that influence which TDP they prefer. Further it is discussed that employing the method of TDPs resulted useful to communicate design choices and to actively involve end-users early in the design process.

Table of content

1 Preface

The work described in this master thesis was performed as part of an internship in the department of Perceptual and Cognitive Systems at TNO[1] which is the Netherlands Organisation for Applied Scientific Research.

Further, the work was performed as part of the European project called ASSISTANCE which stands for "**A**dapted Situation Awarene**SS** tools and ta**I**lored training scenarios for increa**S**ing capabili**T**ies and enh**AN**cing the prote**C**tion of First Respond**E**rs" (Project Nr. 832576)[2]. ASSISTANCE is an international project funded by the Horizon 2020 program in Secure Societies Challenges and it addresses the topic "Technologies for First Responders" (SU-DRS02-2018-2019-2020). The main aim of ASSISTANCE[3] is to help and protect first responders during the mitigation of large disaster and to increase their capabilities when facing complex incidents.

This master thesis project contributes to an early stage of an iterative design process as the design is repeatedly evaluated and refined. The aim is to create a tool that matches the needs of the end-users and is improved in quality and functionality. Therefore, the findings resulting from the work performed in this master thesis will directly feed into the design process.

Firefighters affiliated with the Gezamenlijke Brandweer Rotterdam in the Netherlands have participated in the evaluation of the design work and therefore a special thank you is expressed to them.

This master thesis has led to the publication of a paper in the 4th International Conference on Human Systems Engineering and Design: Future Trends and Applications (IHSED 2021)[4].

[1] https://www.tno.nl/en/about-tno/
[2] https://cordis.europa.eu/project/id/832576
[3] https://assistance-project.eu
[4] Beuker, T., Mioch, T., and Neerincx, M.A. (in press). Team Design Patterns for participatory development of First Response Human-Agent Teaming. In: Proceedings of the 4th International Conference on Human Systems Engineering and Design

2 Introduction

Disasters, whether man-made or of natural origin, pose a serious threat to the lives of citizens, the economy and the environment (Directorate-General for European Civil Protection and Humanitarian Aid Operations (ECHO), 2017). When a disaster occurs first responders (FRs), like firefighters, police officers and paramedics are the first ones to arrive at the scene and deal with the emergency (Cambridge University Press, n.d.). The task of FRs is to provide assistance to whoever is affected, mitigate negative effects and if possible, resolve the cause of the incident. Industrial accidents are described as one of the main disaster risks by the European Union and are mainly addressed in this thesis. Industrial accidents involve the release of dangerous substances, explosions or fire (ECHO, 2017). FRs that deal with these disasters are usually faced with dangerous and complex environments where the information they receive and base their decisions on is uncertain and incomplete. Additionally, these events are usually dynamic and constantly evolving. To make educated decisions about plans and countermeasures FRs have to obtain and maintain adequate situation awareness (Endsley, 1988), meaning that their understanding and mental representation of the problem space should be as accurate as possible given the available information. In addition, they have to constantly update their understanding of the situation and pay attention to dynamic changes in the environment that affect their mitigation efforts. This means that FRs have to recognize when there is a change, understand its impact on the situation and adapt their strategies and actions accordingly to keep the public and themselves safe. Achieving and constantly updating one's situation awareness can be challenging in the dynamically evolving environments FRs work in. The following fictional example[5] illustrates the uncertainty and variability of information that FRs have to base their decisions and actions on.

After receiving an emergency call about an industrial accident, FRs, in this case firefighters, are on their way to the reported incident site. They receive the information that a not yet identified gas is escaping from a factory at an industrial site, but no fire nor victims are reported so far. After calculating the expected trajectory of the escaping gas by considering the wind direction and wind speed, the firefighters conclude that it will spread over an unpopulated area therefore unlikely affecting any citizens. The firefighters decide to not take any special precautions for the moment. This means that they are not sending out an alert to the public, which would inform the citizens about the escaping gas and recommend staying

[5] The example is made-up by the author. It is inspired by interviews that were held and literature that was consulted.

inside and shutting the windows. They are also not yet requesting an ambulance nor additional firefighting teams to assess the concentration of the escaping gas in affected areas. However, once the firefighters arrive at the incident site, they are faced with novel information that requires them to reassess the situation and modify their actions. A factory worker tells the firefighters that the escaping gas is highly toxic and that the cause for the escaping gas is a leak in one of the pipes. Additionally, the wind direction changed, now pushing the gas cloud towards an area where many people live and work. On top of that, another factory worker inhaled some of the escaping gas and is showing symptoms of intoxication. All this new information leads to a reassessment of the situation and to an adaptation of the planned actions. The firefighters decide to send out an alert to the public, to request an ambulance for the injured worker and to call for additional teams that can do measurements in the affected area.

Incidents like the one described in this example require the FRs to constantly evaluate new information and reassess their plans.

Applications and tools that use advanced artificial intelligence (AI) and sensor technology pose a great potential to support FRs in staying resilient, maintaining situation awareness and adapting planned actions throughout dynamic incident developments. It is important that the developed applications and tools complement the capabilities of the FRs, alleviate demanding tasks and do not create additional workload for the FRs (Neerincx et al., 2016).

This master thesis is carried out as part of an internship at TNO[6] and within the EU funded ASSISTANCE project (Project Nr. 832576)[7] which aims to develop such a technologically advanced tool that supports FRs during the mitigation of large disasters. The tool that is being designed will offer a holistic solution that includes different SA supporting applications combined in a wider SA platform (Perez et al., 2020). One module of this SA platform focuses on providing assistance during incidents that involve the exposure of hazardous gas (e.g., caused by industrial accidents). This module is called the Chemical Hazard Tool and it is able to calculate and predict the distribution of an escaping gas cloud. This gas cloud is displayed on a map and is constantly updated through live meteorological data, chemical sensors and input of FRs. The Chemical Hazard Tool indicates whether the public is potentially affected by the gas cloud and therefore in danger. Based on this information FRs are aided in their

[6] https://www.tno.nl/en/
[7] https://cordis.europa.eu/project/id/832576

decision about which measures need to be taken. Additionally, GPS data is used to display the location of FRs on the map and to identify if they are in proximity of potential danger zones (Mioch et al, 2021). The tool and its functionalities are intended to provide the FRs with assistance in noticing, understanding and adapting to dynamic and uncertain developments throughout disaster mitigations. The holistic SA platform that will be the outcome at the end of the ASSISTANCE project will include different applications and it will accommodate the need of several types of FRs. However, for the design of the Chemical Hazard Tool, TNO is collaborating with firefighters and therefore the term FR will be used in this master thesis as a representation for members of the fire brigade.

The work performed for this master thesis contributes to the design process of the Chemical Hazard Tool. We will investigate how the interaction between the tool and the FR should be designed to provide appropriate support during dynamically evolving incidents, assure acceptance by the FRs and meet their needs.

Rather than viewing the designed tool as taking over a task that is usually carried out by the FR, the Chemical Hazard Tool is viewed as a team member that compliments the skills of the FR. The intention is to avoid pitfalls (e.g., prompting workarounds, creating higher workload, requiring more expertise, etc.) which result from disregarding unavoidable interactions with the tool and inevitable changes in the way work is carried out, also described as "substitution myth" (Hollnagel & Woods, 2005). Hence, to anticipate these interactions, the FR and the Chemical Hazard Tool are regarded as team members that coordinate their actions, are mutually dependent, have designated responsibilities and work towards achieving a common goal (Neerincx et al., 2016). This hybrid team constellation is also referred to as Human-Agent Team (HAT), in which the Human represents the FR and the Agent represents the Chemical Hazard Tool. A recently introduced method to design and enhance the collaboration within a HAT are Team Design Patterns (TDPs). These TDPs can explicate core processes and interdependencies within a HAT in a comprehensive and understandable way to a variety of stakeholders, including non-experts in human-agent teaming like end users, project managers or programmers (Van Diggelen & Johnson, 2019). More precisely, TDPs describe roles, responsibilities and adaptive task allocation within a HAT in an abstract and reusable manner, meaning that they are applicable for various situations (Van Diggelen & Johnson, 2019). By using a graphical and textual representation, they facilitate the communication about design choices and thereby provide the possibility to actively involve different stakeholder in the design process (Van der Waa et al., 2020). In this way relevant expert knowledge can be incorporated in the learning and reasoning of the Agent (Van Stijn et al., 2021).

The research question that is intended to be answered in this master thesis results from the aim to design the Chemical Hazard Tool as a valuable team member that supports FRs in staying aware, making adequate decisions and adapting measures when needed throughout dynamically evolving incidents. Working in teams is an intuitive process for people but machines do not have intuition. Therefore, this master thesis will investigate how the Chemical Hazard Tool should act within the HAT and collaborate with the FR. The (explorative) research question is:

> **How should the collaboration between first responders and the Chemical Hazard Tool be designed? Which of the created Team Design Patterns is suitable to support first responders and assures adequate situation awareness?**

To answer the research question, the method of TDPs will be used to design three alternative collaboration options between the members of the HAT. The TDPs will then be compared and evaluated by FRs for their suitability. Because TDPs emerged only recently and research around this topic is still scarce, we decided to assess whether the claims that TDPs are reusable and understandable are also confirmed by end-users, in this case FRs. Therefore, the sub-question is:

> **Do the created TDPs lead to a good understanding of the proposed collaboration designs among the first responders and do the first responders regard the TDPs as reusable?**

To answer the two stated research questions, the work for this master thesis will be split into two parts. In the first part, the possible collaborations within the HAT will be designed using the method of TDPs. In the second part, the proposed design will be evaluated by creating an online survey and distributing it to FRs.

In the first part different TDPs will be created to then identify the best fitting solution within the HAT. Each of the TDPs will vary in roles, responsibilities and tasks assigned to the team members. In the next design step, the TDPs will be transferred into Interaction Design Patterns (IDPs) that describe how the information will be displayed and how the FR can interact with the Chemical Hazard Tool.

For the second part a survey will be created in which FRs are presented with a low fidelity simulation of how the collaboration according to the created TDPs would look like during hypothetical incident scenarios. The suitability of the TDPs will be compared and assessed through different measures. These include preference and situation awareness of the FRs, as well as rated helpfulness, understandability and reusability of the TDPs.

The work performed in this master thesis contributes to the development of the Chemical Hazard Tool and is performed at an early stage of the iterative design process. Building on the findings, the Chemical Hazard Tool can be further refined and tested again in a cyclic process. Further, this master thesis is an addition to the still young and scarce research field around Team Design Patterns.

2.1 Thesis outline

Section 3 provides a description of existing literature on situation awareness, human-agent teaming and design patterns with a specific focus on Team Design Patterns and Interaction Design Patterns. Section 4 describes the design process and the produced TDPs and IDPs. In section 5 the evaluation process for comparing the TDPs and identifying a suitable collaboration style is described. In section 6 obtained results are presented. Section 7 discusses the results in relation to the explorative research questions and provides a recommendation on how the collaboration within the HAT should be designed. Finally, section 8 provides a conclusion, lists limitations and describes further research.

3 Literature Research: Situation Awareness, Human-Agent Teaming and Design Patterns

This section provides an overview of topics that are central to the work performed for this thesis. First, in subsection 3.1 the concept of situation awareness (SA) is explained. SA has been identified as an integral component for performance and safety. It is therefore of great importance for FRs during the mitigation of incidents. Next, Human-Agent Teaming is discussed in subsection 3.2. The concept serves as the basis for designing the Chemical Hazard Tool as a team member. It outlines the importance of a human centred design in which the technology is not replacing the human but complementing the human by acting as a team partner. At last, subsection 3.3 concludes with an overview of Design Patterns that originated in the field of architecture and spread across various fields (e.g. software engineering, interface design, etc.). A special focus is given to Team Design Patterns (TDPs) and Interaction Design Patterns (IDPs) as they are utilized for the collaboration and interaction design between the FR and the Chemical Hazard Tool.

3.1 Situation Awareness

Situation awareness (SA) refers to recognizing critical cues in the environment, understanding what they mean and how they will affect the situation (Endsley, 1988). Having this awareness of what is happening around someone is important when making decisions, as a lack thereof can lead to accidents and failures (Chauvin et al., 2013; Jones & Endsley, 1996). Especially for FRs it is vital to acquire good SA, as they have to make many safety critical decisions while attending missions that pose a threat to civilians and themselves. Uncertain and changing conditions pose a challenge to maintaining SA and require a constant reassessment of one's mental model.

Within their teams each FR should also build and maintain shared SA, meaning that different team members have to be aware of the same key elements in the environment. Shared SA can be established through multiple ways including verbal communication, shared displays and shared environments (Jones & Endsley, 2001). Gorman et al. (2006) emphasize the importance of coordination and communication for Team SA. Every team member holds valuable pieces of information that make up the whole picture. Deciding which information to share and with whom is a substantial team process as unimportant details can be distracting and produce overload but missing essential information can lead to poor decisions. Therefore, delivering the right information at the right time to the right team member is vital for successful team operations (Gorman, Cook and Winner, 2006). Another concept that impacts team performance is mutual awareness. It refers to being aware of what other team members are doing, what they will do next and what their assumptions about the situation are (Shu & Furuta, 2005).

When implementing new systems and tools, SA requirements should always be taken into account, because otherwise these systems can create new sources of errors by leaving the operators out of the loop and diminishing their situation awareness (Endsley & Kaber, 1999). As stated in the introduction, a possible solution is designing and implementing new technologies as team members that share knowledge, pursue team goals and coordinate their actions (Neerincx et al., 2016). Shared SA, mutual awareness and coordination requirements are not exclusively important for human-human teaming but equally for human-agent teaming. The term agent refers to an advanced technology into which commonly artificial intelligence (AI) is built in. These AI agents should be able to establish a shared understanding of the problem, anticipate actions the human will perform and be transparent enough so that the human is able to anticipate the actions of the AI agent. In sum, all parties should participate in coordinating and communicating within the Human-Agent Team (Grigsby, 2018; Schaefer et al., 2017).

3.2 Human-Agent Teaming

After exploring the literature on how to design new systems, tools and applications, two perspectives were identified on how these technologies are assumed to be of benefit for the environments they are deployed in. On the one hand, they can be viewed as a replacement for people by doing the job better, faster and cheaper. On the other hand, they can be seen as collaborators enhancing the capabilities of humans (Bradshaw et al., 2011). This collaboration is also referred to as Human-Agent Teaming (Van der Waa et al., 2021).

The first view was born out of the belief that people have flaws and make errors. Consequently, creating a machine that takes over and performs the same task will lead to lower workload, fewer errors, more efficiency and therefore overall better outcomes. Lists that describe “Man Are Better At – Machines Are Better At” were created to determine which tasks to substitute. However, this type of function allocation disregards that changes in circumstances and time can vary the suitability to carry out a task for either human or machine (Dekker & Woods, 2002). Also, this view is subject to the “substitution myth” because contrary to the expectation that the designed system will only have a small impact on the human’s work, it always changes the nature of the work that needs to be performed and in the end requires the human to interact with the system in some sort even if originally not intended during the designing process (Hollnagel, 2007). Not paying attention to these interdependencies between the human and the machine can lead to increased costs as the job for the human changes, gets more difficult and requires more training and expertise (Blackhurst et al., 2011). Therefore, the second approach namely designing a Human-Agent Team (HAT) and looking at what teaming skills the AI agent should possess is a better approach to create efficient performance and safety at work (Johnson & Vera, 2019).

In a general definition, teams "interact cooperatively and adaptively in pursuit of shared, valued objectives. Further, team members have clearly defined, differentiated roles and responsibilities, hold task relevant knowledge, and are interdependent" (Salas et al., 1993). Due to being interdependent, team members have to organize and coordinate their joint activities. This implies extra work beyond performing the task itself. Hence an AI agent needs the ability for both task-work and teamwork (Bradshaw et al., 2011). Coordination within a team benefits from successful communication, where needed information is anticipated and shared proactively (Demir et al., 2017). Further, as described above, shared and mutual awareness are also crucial for teamwork (Jones & Endsley, 2001; Shu & Furuta, 2005). These demands need to be fulfilled by both the human and the AI agent.

Johnson et al. (2014) propose that to be good teammates, AI agents should be observable, predictable and directable. These requirements are a concise reformulation of the ten challenges for making automation a team player (Klein et al., 2004). Observability is defined as "making pertinent aspects of one's status, as well as one's knowledge of the team, task, and environment observable to others" (Johnson et al., 2014). In other words, it should be clear what the AI agent is doing, when it is changing its behavior and why it is doing it. Predictability is that "one's actions should be predictable enough that others can reasonably rely on them when considering their own actions" (Johnson et al., 2014). It should be possible to anticipate what actions will be performed by the AI agent. If actions taken by an AI agent are not expected or comprehensible to the human teammate, then interaction might be reduced and well-intended features ignored or misused (Marathe et al., 2018). Directability refers to "one's ability to direct the behavior of others and complementarily be directed by others" (Johnson et al., 2014). Because team members are interdependent and strive to achieve an overall goal, it seems obvious that they should also be able to influence and correct each other's actions when it seems like they have gotten off track. In the literature these requirements can be found under different terms. For example, the term transparency includes the requirements of observability and predictability (e.g. Chen et al., 2014).

For good teamwork the AI agent not only needs to be observable, predictable and directable but it also needs to be able to observe, predict and direct the actions performed by the human. This should be made possible by giving the AI agent the capabilities of understanding and monitoring intentions of the humans and the context. In other words teaming should be a design requirement when creating AI agents (Johnson & Vera, 2019). In the section below it will be described how design patterns can be used to effectively integrate these teaming aspects when developing new systems.

3.3 Design Patterns

Design patterns provide reusable solutions for problems that occur repeatedly. They were first introduced by Christopher Alexander (1977) within the domain of architecture. Design patterns are described on an abstract level so that the solution can be implemented over and over again but never looks the same. Therefore, patterns encapsulate the general core of a solution, which then has to be filled in with details depending on the circumstances a concrete problem occurs in. Moreover, patterns should not be looked at in isolation, but as related to each other. They can be combined and be part of a hierarchical structure where smaller patterns are integrated into larger patterns (Alexander et al., 1977). For example, designing a dining room includes smaller patterns like how the lightening should be implemented and the furniture should be distributed. Then again, the dining room itself is embedded in a larger pattern such as a house or an apartment.

Design patterns are a good way to represent and share design knowledge in a structured way (Chung et al., 2004). Therefore, the idea has since been implemented into various fields like software engineering (Gamma et al., 1995), interface design patterns (Van Welie et al., 2001), workflow engineering (Van der Aalst et al., 2003) and interaction design (Borchers, 2001).

3.3.1 Team Design Patterns

With increasing progress in the development of AI systems and shifting towards designing cooperation between AI agents and humans, the pattern approach gained popularity in the field of human-agent teaming (HAT). While humans have the innate capability to learn how to adapt their teaming skills towards new situations and people, AI agents need to be programmed for this purpose. The process of defining and coding teaming skills into an AI agent is not an easy straightforward exercise (Van Diggelen & Johnson, 2019). Although a lot of research has been focusing on describing requirements and guidelines for AI agents as teammates (e.g. Klein et al., 2004), the implementation remains difficult and conventional interaction design methods fail to address the capabilities of artificial intelligent systems adequately. Team Design Patterns (TDPs) attempt to fill this gap and provide the means to capture HAT processes and explore how AI agents can be designed as adequate team members that contribute to team performance, resilience and cohesion (Van Diggelen et al., 2019). A TDP is defined as "a description of generic reusable behaviors of actors for supporting effective and resilient teamwork" (Van Diggelen et al., 2019, p. 16). In Figure 1 an excerpt of a team ontology is presented that shows the properties of a team that can be applied in a HAT. It shows how important elements within teams relate to each other. Team members are actors that can be human or AI agents. Their objective is to reach a goal for which they put a plan into place. This plan can be implicit, decided ad hoc, or planed in advance (Van Diggelen et al., 2019). Besides the benefit of reusing and implementing patterns when designing AI agents as team members for various environments, patterns also provide a base from which evaluations can be made (Shively et al., 2016).

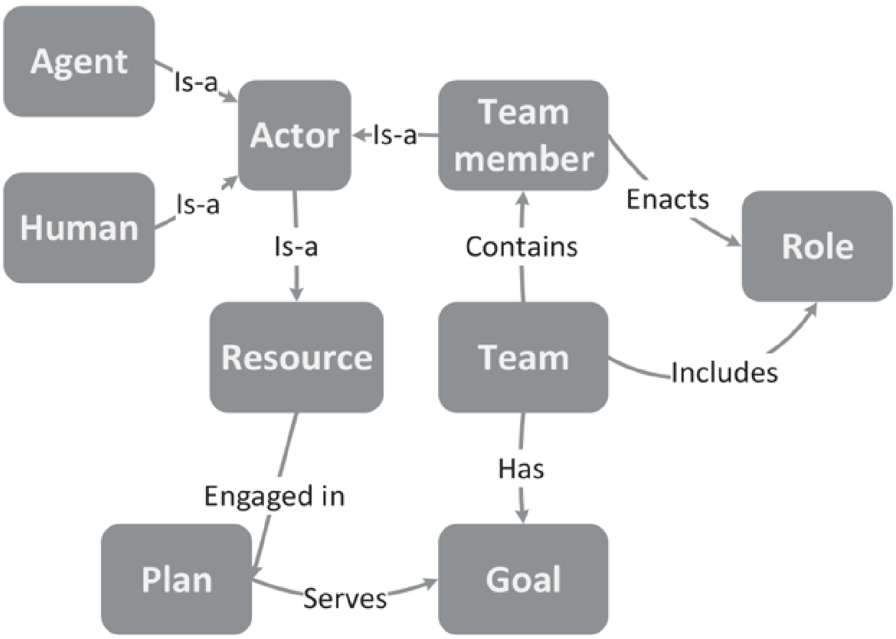


Figure 1. Team Ontology example. Reprinted from Van Diggelen et al., 2019. The figure shows an excerpt of a team ontology that is also applicable for a Human-Agent Team. The ontology shows the relationships between important elements that make up a team. Humans and AI agents can be team members, each fulfils a specific role within the team. Team members are resources that pursue a plan to achieve a goal.

TDPs should fulfil four requirements. (1) They should be intuitive, so that a variety of stakeholders including non-experts understand them and can easily hold a discussion around human-agent teaming solutions. (2) TDPs should be universal enough to include a broad variety of teamwork capabilities. (3) They should be descriptive granting clarity and an accurate distinction between various solutions and situations. (4) And they should be structured so that the intuitive description can be transformed into a more formal specification (Van Diggelen & Johnson, 2019). To fulfil these requirements, van Diggelen & Johnson (2019) propose a simple graphical language that uses icons and symbols to describe the adaptive nature of the work performed in a HAT. Figure 2 shows an example of how human-agent teaming is represented in a TDP. The pattern describes the transition that happens during supervisory control in a HAT, when a handover is performed from the machine to the human and back. Figure 2 shows four small patterns (Human Supervision, Handover to Human, Manual, Handover to Machine) that are nested in the large pattern and comprise it. Due to this simple and intuitive graphical representation TDPs facilitate the communication and discussion about design choices for human-agent teaming (Van der Waa et al., 2020).

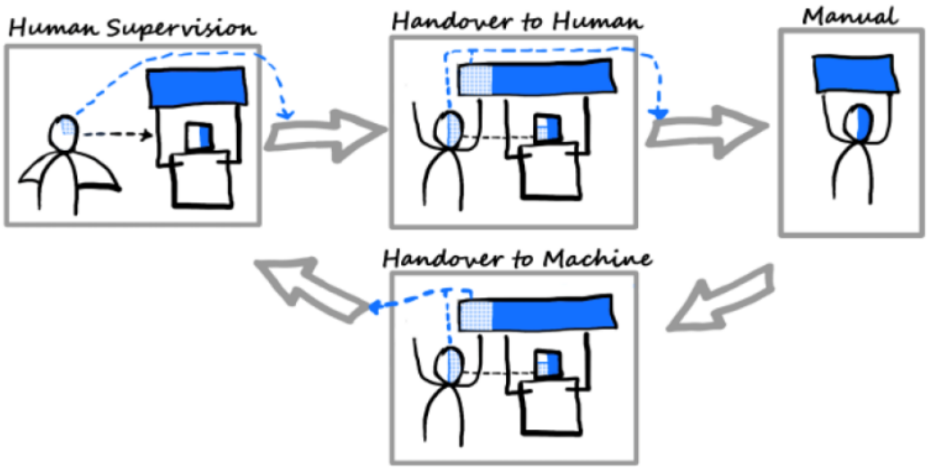


Figure 2. Supervisor control team pattern. Reprinted from van Diggelen and Johnson, 2019. The figure shows how the responsibilities within the HAT change. In the first sub-pattern the machine is performing the work (e.g. self driving car, machine in a factory) and the human is monitoring the task. In the second sub-pattern a handover from the human to the machine happens, this could be initiated for example by an alarm. In the third pattern, the human carries out the task. Once the situation is regarded normal a handover back to the machine is performed. The rounded shaped icons represent the human and the square shaped icons the AI agent. The blocks held above the head represent physical work and the colored heads of the icons represent cognitive work. The arrows indicate the pattern flow.

Additional to the graphical representation, TDPs are described in a written text, usually using a table format. A structured description of a TDP includes:

- a title,
- a description of the collaboration presented in the TDP,
- an example in form of a use case showing how the TDP is implemented,
- requirements that need to be fulfilled by the human and the AI agent,

- advantages and disadvantages of implementing the presented TDP (Van der Waa et al., 2020; Van Diggelen et al., 2019).

TDPs can be generated, linked and nested by using a top-down or bottom-up approach. In the top-down approach, all possible combinations of TDPs are collected and then trimmed according to the scientific literature, evaluations and current technological possibilities. This represents a systematic approach but bears the difficulty of how to define the initial set of TDPs. In the bottom-up approach first a solution for a specific problem within a scenario is found. Then this solution is generalized and evaluated. The benefit is that the emphasis lies on relevant problems and solutions. But on the downside, it is possible that some solutions might not be recognized and therefore no complete solution array would be produced (Van der Waa et al., 2020).

Recent research in the field of TDPs has been focusing on designing and assessing moral decision making within Human-Agent Teams (Van der Waa et al., 2021; Van Stijn et al, 2021).

3.3.2 Interaction Design Patterns

Interaction Design Patterns (IDPs) like TDPs are reusable solutions for recurring problems. Whereas TDPs concentrate on roles, responsibilities and task allocation within a HAT, IDPs describe the look and feel of the designed technology. IDPs are often found in the context of human computer interaction (HCI) and user interface design. The focus is on how information is conveyed and how the user operates or interacts with the system (Borchers, 2001). For example, IDPs could describe how notifications are presented and what options are given to the user to react to those notifications. IDPs, therefore, focus on how the communication within the HAT is designed.

IDPs are usually described in a table format by stating the name of the pattern, the problem it solves, the context in which it occurs, the solution it provides, an example for how it looks like and the rationale behind the suggested solution (Van Welie et al., 2001).

4 Designing Human-Agent Teaming

This section presents the performed design work, in which we utilized the concepts and methods introduced in the preceding literature research. The first subsection describes the design process, including an analysis of the work environment and procedures carried out during incident mitigation by first responders (FRs), more precisely firefighters. Based on this analysis possible joint activities were derived in which the AI agent (Chemical Hazard Tool) could complement the skills of the firefighter. The second subsection presents three Team Design Patterns (TDPs) that were created during the second step of the design process. The created TDPs vary in roles, responsibilities and tasks assigned within the Human-Agent Team (HAT). The last subsection describes Interaction Design Patterns (IDPs) which were created during the third design step, in which the abstract TDPs were specified into IDPs. The IDPs describe how the interaction between the AI agent and the FR is represented on the interface, meaning for example how new information is displayed and what features are provided to react to this information.

4.1 Design Process

In human-human teams as well as human-agent teams it is essential that team members coordinate their actions as they rely and depend on each other. This means that there has to be a consensus about what tasks need to be carried out and who will carry them out. When intending to design an AI agent as an effective team member it is therefore important to identify the tasks or joint activities that can be carried out within the HAT and determine how their coordination looks like (Bradshaw et al., 2011; Klein et al., 2005). For this reason, the first step in the design process was to investigate and understand the nature of the work that FRs, in this case firefighters, carry out. This included an analysis of the environment they work in, procedures they follow, the equipment they use, measures they put in place and roles and responsibilities they hold. A document analysis was performed using information sources on procedures and training materials available online[8,9,10] and by reviewing four documented workshops that were carried out before the start of this master thesis project[11]. Additionally,

[8] Scenarioboek Externe Veiligheid. Accessed: July 18, 2021. [Online]. Available: https://www.scenarioboekev.nl/

[9] Brandweeracademie. Brandweer Nederland. (2015). *Branchestandaarden blijvende vakbekwaamheid*. Accessed: July 18, 2021. [Online]. Available: https://www.ifv.nl/kennisplein/kwaliteit-brandweerpersoneel/publicaties/branchestandaarden-blijvende-vakbekwaamheid

[10] Brandweer Kennis Net. Brandweer Flevoland. (2014). Accessed: July 18, 2021 [online]. Available: https://www.ifv.nl/kennisplein/cbrn/publicaties/handboek-meetplanleider

[11] Workshops performed with the Gezamenlijke brandweer Rotterdam on three occasions (23,29, 30 July 2019). Each time the workshop was performed with 5 participants.

two interviews were conducted with a firefighter from the fire brigade in Rotterdam. This gave further insight into the procedures during disaster mitigation and the expectations on how the Chemical Hazard Tool could be useful during a mission. As a result of the performed work analysis, a concept map was created that can be viewed in Appendix A. The concept map depicts key concepts (e.g., actions, objects, information, tasks, decisions, etc.) and their relations during the mitigation of an emergency. Boxes are used to represent the collected information about concepts. Labelled arrows connect the boxes and indicate their relationship by using linking phrases like “depends on”, “includes”, “needs” or “leads to” (Novak & Cañas, 2008).

Additional to the concept map, a table was created that describes decisions usually taken during the mitigation of an emergency and associated required information. This table can be viewed in Appendix B.

At the end of the first step in the design process, we were able to identify joint activities that could be performed within the HAT and therefore decisions that the AI agent could assist the firefighters with. The identified joint activities included the assessment of the situation and derived action planning. In concrete, the FR and Chemical Hazard Tool could jointly assess the situation. The FR assess the situation by monitoring the environment (including weather conditions) and actively exploring the incident site. The Chemical Hazard Tool assesses the situation by receiving input from live meteorological data, chemical sensors, GPS and direct input from the FR. Another joint activity that was identified is action planning, which builds on situation assessment and involves mission relevant decisions. Deriving from these two generic joint activities four specific tasks (later also referred to as decisions) that could be carried out in collaboration within the HAT were defined. During the evaluation phase these four tasks were implemented in the low fidelity simulation and presented to the FRs. The tasks are:

- Choosing a safe route from the fire department to the incident side. A safe route avoids passing through the trajectory of a spreading gas. For this task the meteorological data needs to be assessed and the gas cloud trajectory calculated. Based on that the route can be chosen.

- Deciding whether to alert the public about an escaping gas. This decision depends on the toxicity of an escaping gas and whether vulnerable locations[12] are detected within its trajectory.
- Instructing measurement teams where to do measurements. When calculating a gas cloud distribution an inherent uncertainty is always present. To reduce this uncertainty, measurement teams are usually send out to measure gas concentrations within the affected area. Based on these measurements, the calculation of a gas cloud distribution can be adjusted and made more accurate. Deciding where to send these measurement teams is done strategically dependent on the predicted gas cloud trajectory and other factors, like vulnerable locations.
- Deciding whether to warn FRs that are located close to or within a danger zone. FRs in the field can find themselves in danger areas in which they can be exposed to the toxic gas. By monitoring the FRs in the field through GPS and assessing the gas cloud distribution, they can be warned if needed.

Section 5.2 further explains the setting in which these tasks were assessed.

In step 2, Team Design Patterns (TDPs) that assign roles, responsibilities and tasks within a HAT were created. These TDPs resulted from evaluating how the above tasks could be completed within the HAT. Therefore, at the core of step 2 was the design of possible collaboration styles between the FR and the Chemical Hazard Tool. As a characteristic of Design Patterns is that they are reusable solutions for reoccurring problems, it was first assessed whether already existing TDPs could be identified within the literature and if they could be reused or adapted for the HAT in this master thesis. In the end, three TDPs were created that represent increasing levels of responsibility and decision support offered by the AI agent (Chemical Hazard Tool). Section 4.2 explains in detail the three created TDPs.

In step 3 Interaction Design Patterns (IDPs) were created. They describe how information is displayed on the interface and what options the FR has to introduce information or access functionalities (e.g., notifications, buttons to accept or decline, etc.). This was done by again consulting the literature for existing IDPs and by anticipating interaction needs that would arise when solving incidents and collaborating within the HAT. Additionally, once the IDPs were drafted onto the interface of the Chemical Hazard Tool they were shown to an interaction

[12] Vulnerable Locations refer to buildings and areas where people are at risk to be affected by the impact of the incident and may have difficulties in getting to safety. This includes for example hospitals, nursing homes and areas of dense populations (Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. VROM, 2010)

designer to identify and correct issues at an early stage. Section 4.3 describes in detailed the created IDPs and according interface of the Chemical Hazard Tool.

4.2 Team Design Patterns for the Human-Agent Team

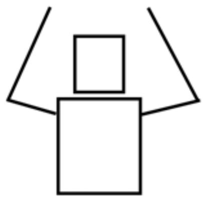
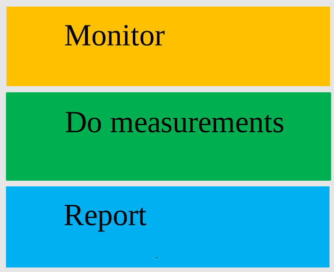
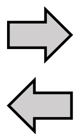
This subsection describes the created TDPs within step 2 of the design process and it provides an explanation of their graphical representation. After identifying what tasks could be performed within the HAT, the TDPs were designed to describe how the collaboration could look like while solving the tasks. As Design Patterns are reusable solutions for reoccurring problems, we searched within the literature for TDPs that could be reused for the purpose of designing the collaboration between the Chemical Hazard Tool and the FR. The second and third TDP were based on previous work done by Van Stijn (2020) and Van der Waa et al. (2020). The TDPs they propose were created within the framework of moral decision making in the medical settings and were accordingly adapted and modified to match the problem space of incident mitigations that involve the escape of toxic gases. The first TDP was created as a logical precursor because the AI agents described within the three TDPs increase in their degree of capabilities and responsibilities from one TDP to the other.

All three created TDPs are aimed at supporting the FR in understanding dynamic incident developments and in employing adequate measures but they differ in roles, responsibilities and tasks assigned to the AI agent and the FR. The AI agents described within the three TDPs represent different options that can be programmed into the Chemical Hazard Tool. The capabilities and responsibilities of these AI agents increase gradually from TDP1 to TDP3. In the first TDP the FR collaborates with an *Informing Agent*. The *Informing Agent* is responsible for updating the FR about the current and predicted situation. In the second TDP the FR collaborates with an *Advising Agent.* Like the *Informing Agent*, the *Advising Agent* is responsible for providing updated information but in addition, it is also responsible for giving mission-relevant recommendations. The third TDP describes the cooperation with a *Deciding Agent*. This *Deciding Agent* is able to make autonomous decisions about mission relevant actions and carry them out.

In all three TDPs, the AI agents base their support on information that is available to them about meteorological data (wind direction, wind speed, weather forecast), GPS locations of FRs in the field, measurements of gas concentrations in affected areas and vulnerable locations. This information is used by the AI agents to visualize and calculate an expected gas cloud distribution.

The TDPs are described both textually and through graphical representation using icons and symbols. Table 1 provides an overview and explanation of the icons and symbols used for the design of the three TDPs presented in this master thesis. The icons and symbols have been adopted and modified from previous work done by Van Stijn (2020) and Van der Waa et al. (2020). Which TDPs were adapted is described below.

Table 1. Explanation of graphical language used for TDPs.

Concept	**Type**	**Explanation**	**Symbol**
Actor	Human	This icon represents the human actor. The arms of the huma actor are raised when s/he holds up a block and therefore performs an activity.	
	AI agent	This icon represents the AI agent. The arms of the AI agent are raised when it holds up a block and therefore performs an activity.	
Work		The block shows information about the task or activity that is carried out. E.g. the human actor does measurements in an affected area to determine the concentration of a toxic gas. When more than one activity or task are carried out the blocks receive multiple colors.	
Transition	Automatic	The solid arrows indicate a transition between frames. From on to the other assigned roles, responsibilities and tasks change.	

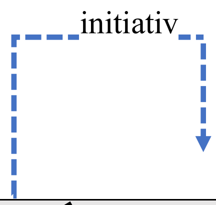
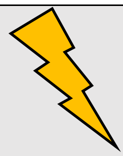
	Initiative	The blue dotted line indicates who initiates a transition from one sub-TDP to the next.	
Change		The lightning icon indicates new information, which affects the incident and requires a reassessment of the situation. An example is a change in wind direction and wind speed.	

Figure 3 gives an example of how the elements shown in Table 1 are integrated within the TDPs. In this master thesis all created TDPs consists of two sub-TDPs, each presented within a frame. Actors, human and AI agent, are depicted within these frames. In the figure the blocks that the actors hold up indicate tasks that are allocated to them. Blocks, i.e. tasks, that are held up together by both actors are performed jointly. Dynamic task allocation is represented by moving from one frame to another following the solid arrows. A transition between TDPs can be initiated by one of the actors, which is indicated by a dashed arrow. The initiation of a transition can be triggered for example by one of the actors receiving new information (e.g. wind direction change), that requires a reassessment of the situation and possibly an adjustment of employed measures. New information is indicated in the figure with the lightning icon. Examples of new information that would require a reassessment of the situation during the mitigation of incidents are updates on weather conditions, deviating gas concentrations measured in affected areas, or issues and delays that might arise while working on resolving the cause for the escaping gas. When a reassessment of a situation is needed, FRs have to consider how the new situation affects their mitigation intentions (e.g. need for extra resources), and the safety of the public and themselves. Once a change has been noticed by one of the actors, succession to the next frame is initiated, and when the change has been dealt with, the collaboration transitions back to the initial frame. We added to the pre-existing list of symbols and icons found in literature, the lightening icon and different colors for the blocks to make the variety of actions more visible (Van Diggelen & Johnson, 20019; Van Stijn, 2020).

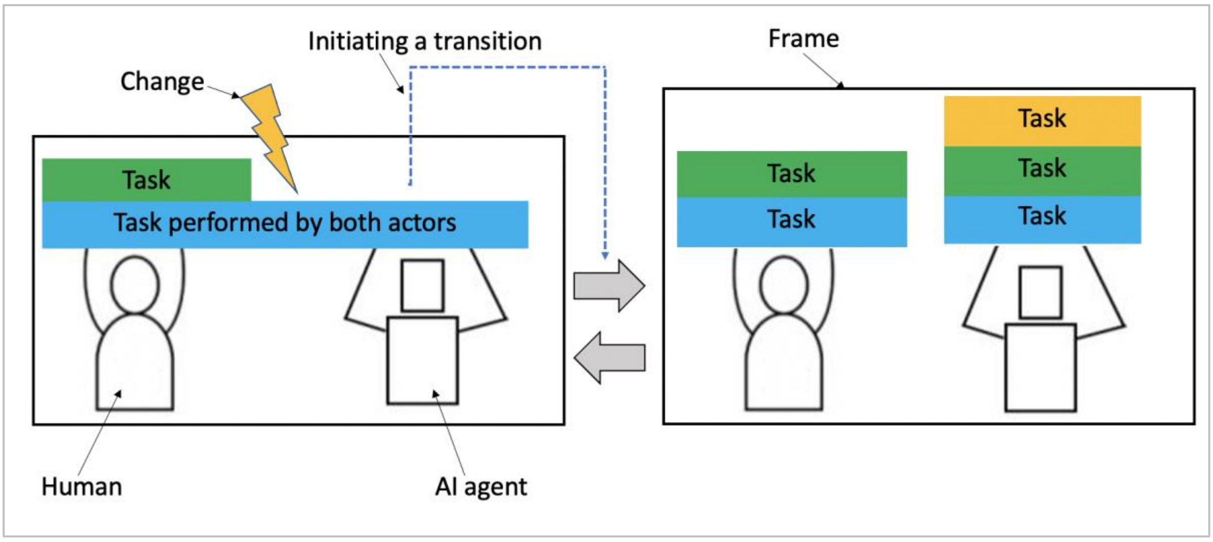


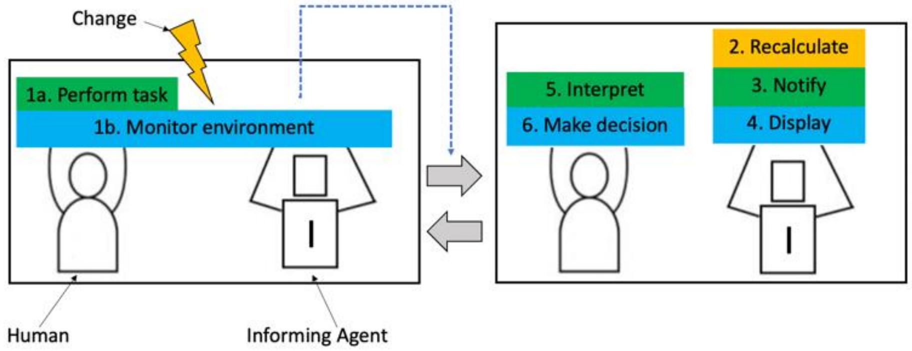
Figure 3. Example of the structure of a TDP.

4.2.1 TDP 1: Informing Agent

In TDP 1, the *Informing Agent* has the responsibility to track the environment and inform the human actor about the current state of the situation and any changes that it might register. In the first frame both the human and the AI agent monitor the environment. Additionally, the human performs a task (e.g. closing a leak). When the AI agent registers a change in the environment, it initiates a transition to the next frame and performs a recalculation of the situation and expected gas cloud distribution. Then it notifies the human actor and displays the adapted situation. The human then is responsible for interpreting the new circumstances of the situation, for thinking of possible ways to adapt to the new situation and for deciding on one of the options. Table 2 provides a detailed description of TDP 1 including the graphical representation. TDP 1 relies on the human having sufficient cognitive capacity to deal with the workload. Overload might lead to reduced situation awareness and decision-making performance (Berggren et al., 2011).

Table 2. TDP 1: Informing Agent.

Name:	**Informing Agent**
Description:	Both actors are monitoring the environment and the human actor additionally performs a mission related task when new information about changes in the environment emerge. The AI agent recognizes the change (e.g. through sensor input) and initiates a recalculation of the model. It then notifies the human actor and

	displays the adjusted situation (e.g. gas cloud distribution). The human actor interprets the implications of the new situation, considers possible actions and makes a decision.
Structure:	
Example	The commander is giving instructions regarding the next steps towards resolving the source of danger, for example giving instructions on how to close a leak, when the AI agent receives updated sensor data that indicates the need for a recalculation of the gas cloud distribution. The AI agent calculates the new model, notifies the commander and displays the new gas cloud distribution. The commander examines the updated model and decides whether and how to adapt measures (e.g. if a region needs to be warned; if firefighters should be sent to other areas to perform gas measurements; if other resources or back up is needed).
Requirements	**R1** The *AI Agent* is able to monitor the environment. **R2** The *AI Agent* is able to automatically update the model when new information arises. **R3** The *AI Agent* is able to display and communicate changes in the environment in an understandable and coherent way to the *Human Actor*. **R4** The *Human Actor* needs the capacity to process and prioritize all relevant information. **R5** The *Human Actor* alone is responsible for generating ideas for possible actions and decides which measures to implement and therefore needs the capacity to perform these activities.

Advantages	**A1** The *Artificial Agent* is not required to interpret changes in the environment with regard to actions that should be taken. **A2** The *Human Actor* stays in control of all mission relevant decisions.
Disadvantages	**D1** The *Human Actor* might experience an overload and therefore miss important information.

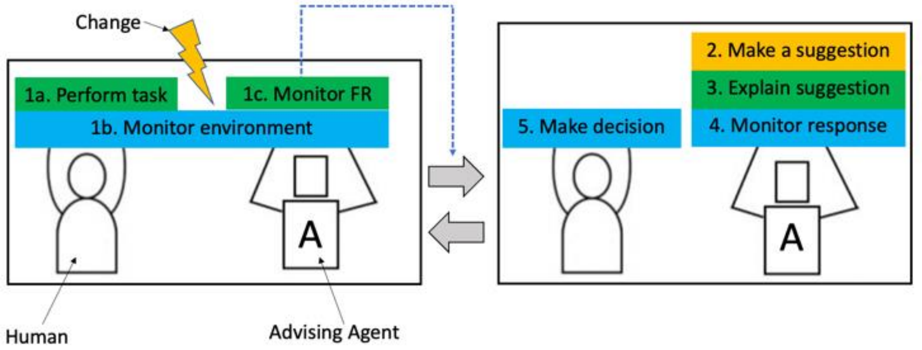
4.2.2 TDP 2: Advising Agent

TDP 2 is similar to TDP 1, but in addition to providing information about the situation, the *Advising Agent* also gives advice on what measures to take or adjust when dealing with an incident that involves the spread of hazardous gas. TDP 2 is an adapted version of the *Suggesting Machine*, which is a TDP proposed by Van Stijn (2020) within the scope of bias mitigation in the medical domain. In TDP 2, the *Advising Agent* provides recommendations with respect to what actions are regarded as most suitable by the AI agent as a reaction to mission developments and changes in the environment (e.g., a recommendation could be to alert the public). Additionally, the *Advising Agent* provides an explanation regarding the specific advice it has given. Offering an explanation supports comprehending the reasoning behind the advice and makes it easier for the human to decide about its accuracy. Therefore, an AI agent that explains itself has increased predictability (Doran et al., 2017; Van Diggelen et al., 2019). Once an advice is given the human actor can decide to follow it or decline it and introduce a different course of action.

Whereas in TDP 1 the human actor was solely responsible for generating possible options for which actions to take during the dynamically evolving mission, in TDP 2 the *Advising Agent* alleviates this workload by stepping in and considering possible options and recommending the most appropriate one. The human is still in charge of making the decision, but the information is presented in a way that it can be immediately acted upon. Table 3 provides a detailed description of the collaboration between the FR and the *Advising Agent*.

Table 3. TDP2: Advising Agent.

Name:	**Advising Agent**
Description	Both actors are monitoring the environment. The human actor additionally performs a mission related task, and the AI agent monitors the human. New mission related information emerges.

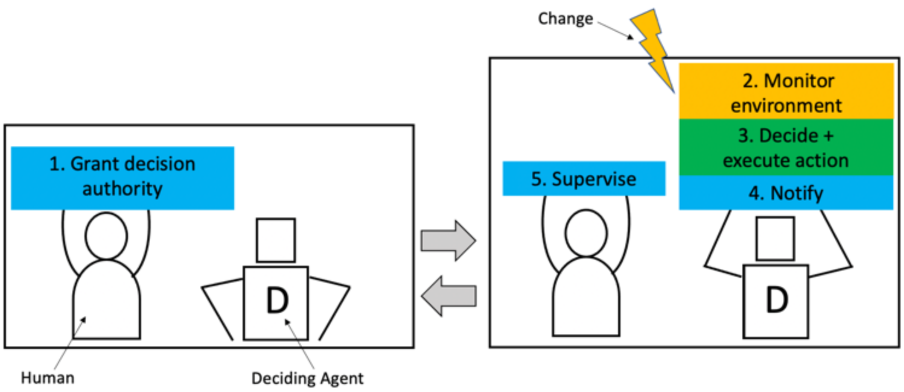
	The AI agent recognizes the change (e.g. through sensor input) and initiate a recalculation of the model and its implications. Based on this it generates an advice about which actions to take, explains it to the human actor (additionally to displaying the new situation) and monitors the response of the human actor. The human actor decides to accept or decline the advice.
Structure	
Example	Measurements of gas concentrations lead to a recalculation of the gas cloud distribution. The AI agent displays the adjusted gas cloud distribution and recommends further measurement locations to increase the certainty about the gas cloud trajectory. The FR accepts or declines (or ignores) given advice.
Requirements	**R1** The *Human Actor* needs to have sufficient understanding of what is happening in the environment. **R2** The *AI Agent* has to understand the implications of the change and provide a suggestion for actions. **R3** The *AI Agent* has to be able to explain what changes happened and why an advice is given. **R4** The *AI Agent* has to be sufficiently trusted in its ability to give advice.
Advantages	**A1** The *Human Actor* is actively supported in the decision-making process. **A2** The *AI Agent* does not need to understand the implications of the proposed action, only the implications of the change that happened.
Disadvantages	**D1** Constant suggestions might annoy the *Human Actor*.

	D2 Not well calibrated advice might confuse and distract the *Human Actor*.
	D3 The *AI Agent* can only produce predesigned advice.

4.2.3 TDP 3: Deciding Agent

TDP 3 describes the *Deciding Agent.* In this TDP the AI agent is allowed to decide on its own and immediately act upon those decisions. This pattern is adapted and modified from the *Autonomous Moral Decision Making* pattern presented by Van der Waa et al. (2020). In TDP 3 the human actor grants the AI agent decision authority for specific types of decisions. Then the *Deciding Agent* autonomously monitors the environment, reassess the situation when needed, decides on an action and performs it. Further, once a decision and action have been performed the *Deciding Agent* notifies the human actor. If the chosen decision is not adequate, the decision authority for future similar tasks can be revoked by the human actor. Being able to revoke decision authority assures a degree of directability of the *Deciding Agent*.

Table 4. TDP3: Deciding Agent

<table>
<tr><th>Name</th><th>Deciding Agent</th></tr>
<tr><td>Description</td><td>The human actor grants decision authority for a specific scope of tasks. The AI agent has then the responsibility to monitor the environment and when it recognizes a change, it makes a decision within the granted scope about which actions to take. Afterwards it notifies the human actor about the performed actions.</td></tr>
<tr><td>Structure</td><td>
</td></tr>
<tr><td>Example</td><td>The AI agent is granted permission to independently reallocate measurement teams in the field. A change in wind direction prompts a</td></tr>
</table>

	recalculation of the gas cloud distribution whereupon the AI agent assigns the teams to new measurement locations. The commander is notified about the new gas cloud distribution and the reallocation of measurement teams.
Requirements	**R1** The *AI Agent* has to be able to understand and predict the implications of the actions it puts in place. **R2** The *Human Agent* has to trust the *AI agent* to be able to make adequate decisions.
Advantages	**A1** Hypothetical reduction of workload for *Human Actor*. **A2** Rapid action implementation.
Disadvantages	**D1** The variability of incidents that FRs are called to might cause mistakes in decision making of the *AI Agent*.

4.3 Interaction Design Patterns for the Human-Agent Team

After analyzing the work of FRs during the mitigation of chemical incidents, defining tasks that can be carried out jointly within the HAT and identifying three TDPs that describe different ways the FR and Chemical Hazard Tool can collaborate while solving tasks, this section concentrates on the outcome of the third design step which is the design of the interface for the Chemical Hazard Tool. The focus is on the look and feel of the interaction between the FR and the Chemical Hazard Tool. For this purpose, the method of Interaction Design Patterns (IDPs) was employed, for which interaction requirements were identified by envisioning how the formulated tasks would be solved within the HAT while collaborating according to the different TDPs.

As a first prototype of the Chemical Hazard Tool already existed when the work for this master thesis was performed, many interaction requirements were already covered by that prototype. Below first the interface of this initial prototype is described. Then an overview is given of interaction requirements that were not covered by the old interface and that lead to the creation of new visual elements and functionalities. The last part describes three concrete IDPs between the FR and the Chemical Hazard Tool, their anticipated interaction problem and according solution. This last part is annotated in the usual form of IDPs (Van Welie et al., 2001).

4.3.1 Original Chemical Hazard Tool Interface

The interface of the Chemical Hazard Tool before the redesign can be seen in Figure 4. The main element is a map on which the calculated dispersion of the gas cloud is visualized. A sidebar is provided on the left of the display. There the FRs can fill in information about an escaping gas and request a calculation of the expected gas cloud distribution. The FRs can choose between different visualizations of the gas cloud distribution, including template[13], contour[14] or ensemble[15]. For the evaluation of the Chemical Hazard Tool only the template and contour visualization were used. In Figure 4 a template visualization is shown and in Table 6 a contour visualization is shown. Further information that the FRs can indicate are the location of the release source, the time when release started, the chemical that is being released, the estimation of release rate and released amount. Wind speed, wind direction and atmospheric stability are obtained automatically by the Chemical Hazard Tool using meteorological data. The FRs have the option to overwrite this automatically obtained information and adjust it as they see necessary. FRs often deal with situations where they do not have all the information available, therefore it is also enough to only indicate the location of the accident and the Chemical Hazard Tool calculates a rough estimation of the gas cloud distribution.

For more information on the Chemical Hazard Tool, it’s functionalities and the underlying model on which the gas cloud distribution is calculated Mioch et al. (2021) can be consulted.

[13] Template visualization: Gives a rough indication of where the hazardous cloud is likely to go. Figure 4 shows a Template visualization.

[14] Contour visualization: Gives a more accurate impression by showing the calculated width and length of the gas plume. Additionally, three toxicity levels (life-threatening, potentially harmful and discernible) are indicated by three lines within the gas plume (see picture in Table 6).

[15] Ensemble visualization: Shows a set of gas plumes that depict the most likely trajectory of the gas plume and its possible deviations. This indicates uncertainties of the calculation.

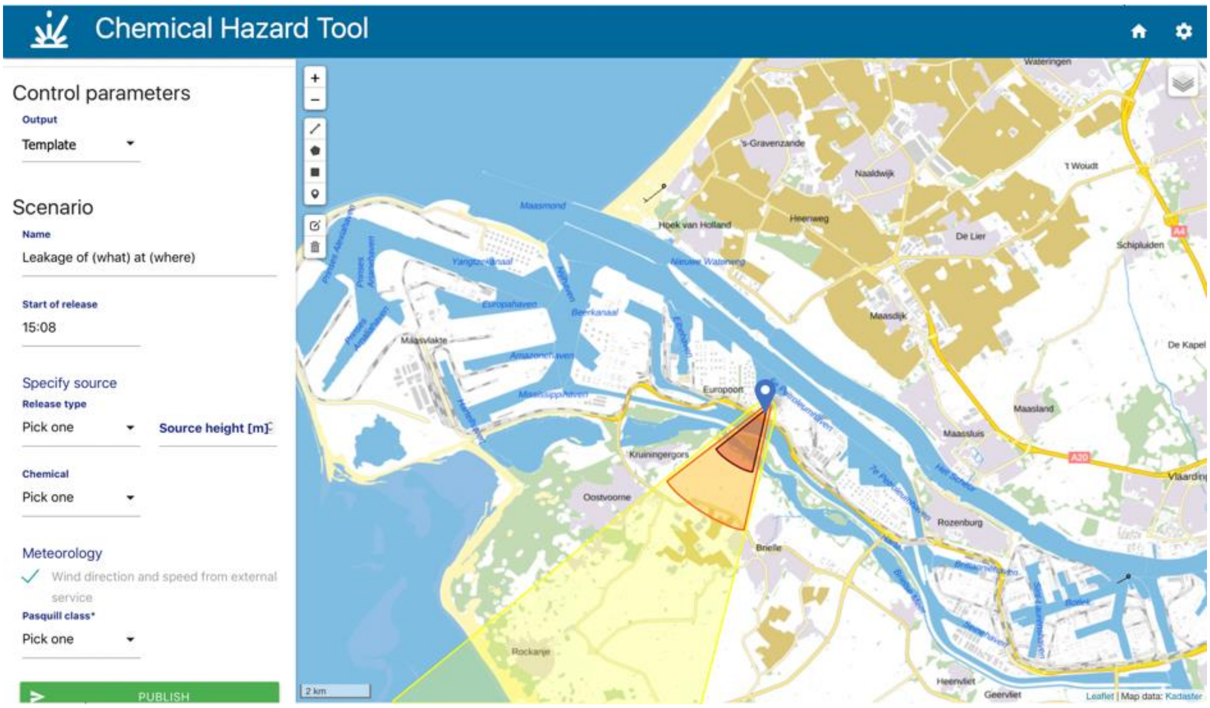


Figure 4. Interface of Chemical Hazard Tool before modifying it to the interaction requirements derived from created Team Design Patterns (TDPs).

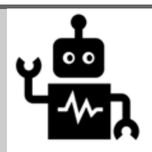
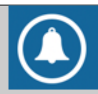
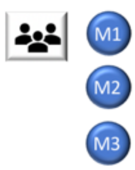
4.3.2 Additional Icons and Functionalities for the Chemical Hazard Tool

This section presents the icons and functionalities that were added to the interface of the Chemical Hazard Tool. Interaction requirements were identified by envisioning how the aforementioned tasks would be solved within the HAT. Thereof we derived what information should be visualized on the interface. We created new visual elements and functions that would convey this information and had them reviewed by an interaction designer at TNO. Table 5 presents the icons and according functionalities that were added to the interface at the end of this process.

The requirements that were identified and had to be addressed within the interface were:

- providing new information and explanations about registered changes (e.g., wind direction change) and adapted gas cloud distribution;
- the opportunity to track FRs in the field via GPS;
- indicating recommended measurement positions to identify the concentration of the hazardous chemical in an area and to save these measurements within the tool;
- providing an indication of vulnerable locations;
- and providing a reference of the AI agent that collaborates with the FR.

Table 5. List of icons added to the chemical hazard module interface.

Description	**Icon**
Portraying the AI agent. This icon has the purpose to create a reference point for when the AI agent provides new information to the FR.	
This icon indicates new information. For example, when indicating that the gas cloud distribution was recalculated.	
This icon opens and closes the notification panel.	
These buttons indicate available active measurement teams and allow directly assigning them to specific measurement positions.	
This icon indicates the GPS location of a measurement team. The number in the circle refers to the specific measurement team.	
The white drop pin icon shows already measured positions. When the FR clicks or hovers over the icon the measured value is supposed to be displayed.	
The black drop pin icon indicates assigned measurement locations that are not yet measured. The number within the pin refers to the designated team that is assigned to do the measurements.	
The icons on the right represent vulnerable locations. The icons that were used within the evaluation are shown on the right. This list is not exhaustive and has to be extended according to the variety of existing vulnerable locations. Icon 1 represents a highly populated area, icon 2 represents a hospital, the caravan indicates a camping area, icon 4 represents a prison and icon 5 represents a nursing home.	1.  2.  3.  4. 

	5. 

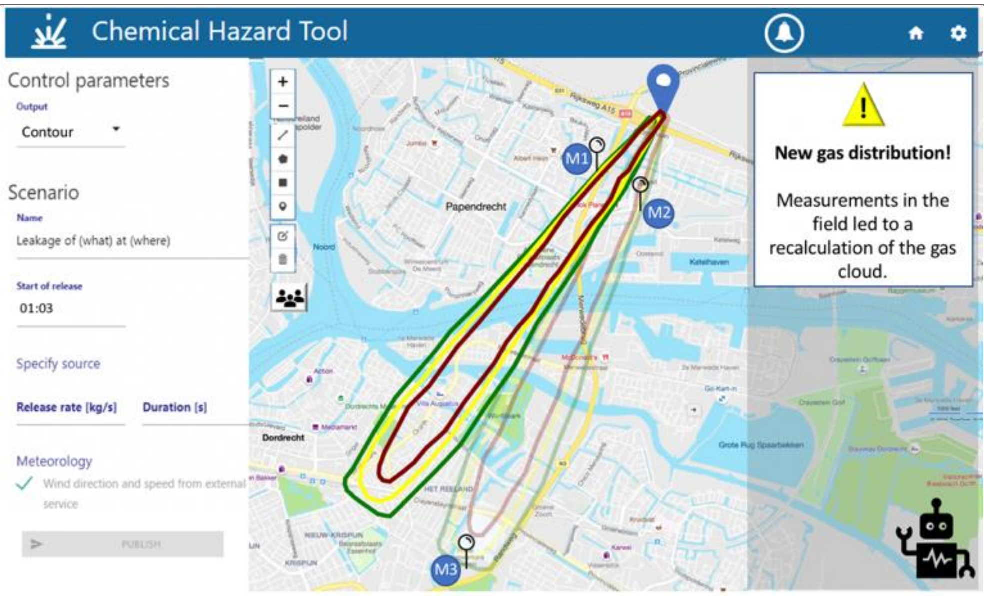
Further interaction requirements and their adaptation within the interface are described in the following three sub-sections.

4.3.3 IDP 1: Communicate Change

The chemical hazard module receives updates from measurements in the field and live meteorological data. When a change is registered that leads to updating the gas cloud distribution it has to be assured that the FR notices the difference between the previous and updated status. The created solution in this pattern is that the old cloud distribution is still visible but faded. Additionally, a notification is given in the notification panel on the right. Table 6 describes this IDP in more detail.

Table 6. IDP 1: Communicating change in the environment and its consequences on the situation.

Name	**Display/Communicate Change**
Problem	The displayed visualization is updated, and it has to be made sure that the human actor recognizes this change.
Context	The user has to make an assessment of the changed circumstances and it is important to be able to compare the previous state with the new state.
Solution	To increase the observability of the change, the old status should be displayed in a way that makes its outdated status clear, for example by fading it out or using dotted lines. The new status should be displayed in full color. Additionally, an explanation to what has happened and why the visualized status was updated needs to be provided.
Example	In this example below the old gas cloud distribution is displayed in faded colors. On the top right a notification is provided that directs the user's attention to the updated situation and that explains the reasons behind the recalculation.



Design rationale	Showing both, the old and new gas cloud distribution assures that the user can make accurate assumptions about the extent of the registered change. The explanation in the notification panel directs the user to the change and additionally provides valuable context.

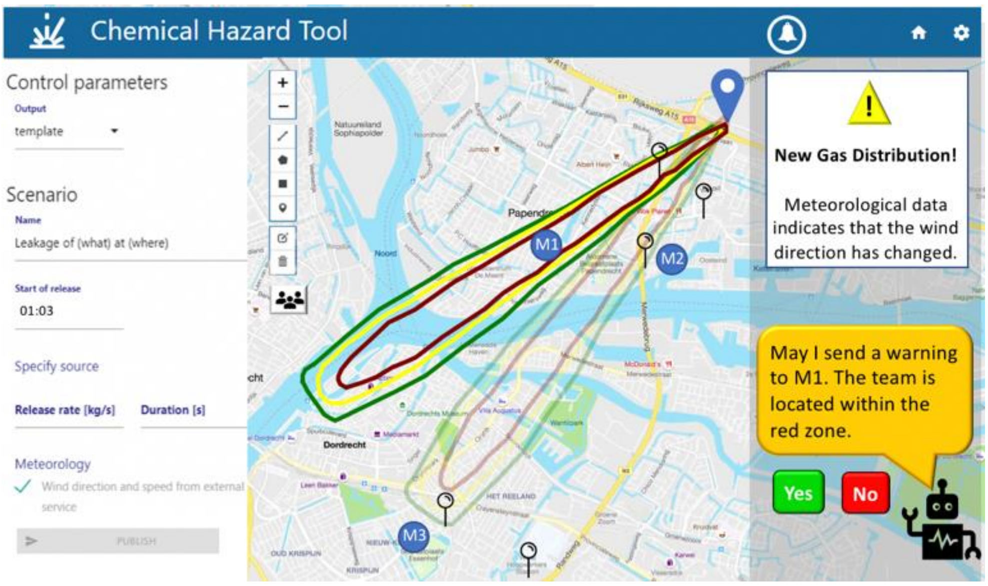
4.3.4 IDP 2: React to Recommendations from the Advising Agent

TDP 2 introduces an AI agent that is able to give advice about which actions to take during missions. Here the according IDP is presented. It is described how the advice given by the AI agent is presented and how the FR can react to this advice. In the example below, the AI agent registers a change in wind direction. After recalculating the gas cloud, it detects that one measurement team (M1) is located within the danger zone and suggests to warn that team. FRs carry smartphones with them that are directly linked to the AI agent. Therefore, the AI agent can directly send a warning to the affected teams. The FR can decide to accept or decline the advice of warning the team members. The FR can indicate the decision by clicking the “yes” or “no” button.

Table 7. IDP2: Present a recommendation and provide the option to react to it

Name	**React to advice**
Problem	The user should receive enough information to weigh up the advice and communicating the decision to the AI agent should be simple.

Context	The AI agent suggests a measure and the human actor needs to understand the suggestion and make a decision about its legitimacy.
Solution	Provide information about the circumstance that led to the advice and implement buttons with which an action can be accepted or declined.
Example	In the example below the AI agent suggests warning measurement team 1 (M1). This suggestion from the AI agent is provided in the yellow bubble and the accompanying explanation is provided in the white box on the top right. Additional to the textual explanation the human actor can also consult the map to verify the information. The "yes" and "no" button provide the opportunity for the user to react to the advice.

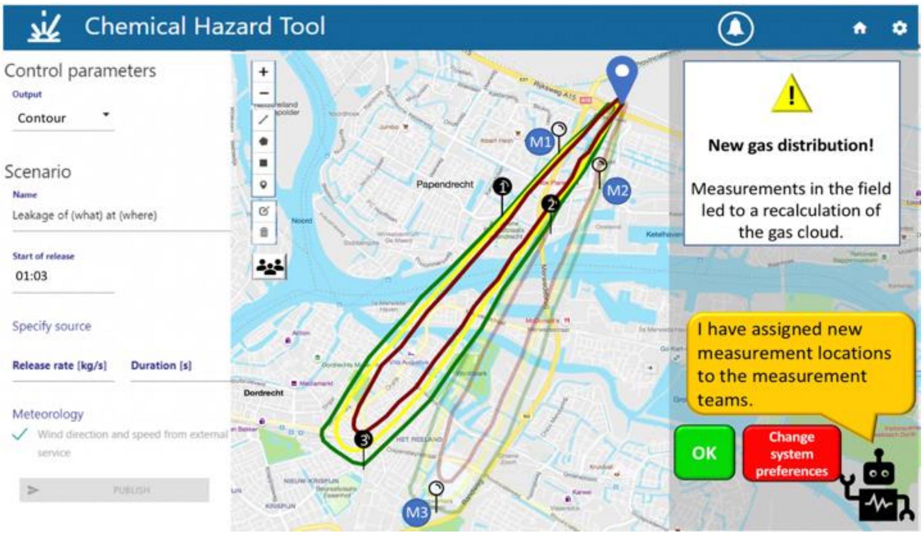


Design rationale	Explaining the reasoning behind given advice makes the AI agent more predictable and observable. The provided option to accept or decline the advice makes the AI agent directable.

4.3.5 IDP 3: Assure directability of *Deciding Agent*

In the third TDP the *Deciding Agent* is given permission by the FR to decide on its own and act upon these decision in the beginning of the mission. The pattern also denotes that the *Deciding Agent* needs to inform the FR about actions that were initiated. Further it should be possible for the FR to take back the right from the *Deciding Agent* to perform certain decisions. How this interaction is implemented is described in Table 8..

Table 8. IDP 3: Assuring the directability of the Deciding Agent.

Name	**Assure directability of Deciding Agent**
Problem	A user might want to withdraw the decision authority that was granted to the AI agent.
Context	The AI agent performs an action that the user does not agree with. The user wants to disable the option that the AI agent keeps making decisions by itself.
Solution	After the AI agent has made a decision and acted upon it, it informs the user. Then the user can acknowledge it by clicking "OK" or if the action is not in line with what the user considers correct the decision authority can be taken back by clicking "Change system preference". If the user clicks on the second option, the choice is given to change the settings to either TDP 1 or TDP 2 for this type of decisions.
Example	Teams in the field do measurements to reduce the uncertainty of the gas cloud distribution. The AI agent keeps assigning new measurement locations to the measurement teams and informs the FR about these actions. If the FR does not want the AI agent to continue with assigning measurement locations to teams, the FR can decide to click on the "Change system preference" button and therefore disable the *Deciding Agent*. 
Design rationale	Giving the option to revert the granted decision authority makes the system more directable and adaptable to the needs of the human actor.

4.3.6 Adapted Interface and Changes in Interaction According to TDPs

Figure 5 shows how the three AI agents vary when interacting with the FR. In the example the different AI agents assist the FR with deciding whether to alert the public about the escaping gas. The *Informing Agent* only shows the template gas cloud distribution and depicts the vulnerable locations. The *Advising Agent* suggests to send an alert to the affected area. And the *Deciding Agent* only informs the FR after it has already sent an alert to the area affected by the calculated gas cloud distribution. Appendix K shows screenshots of interfaces conveying how the Chemical Hazard Tool interacts with the FR according to the task at hand and the AI agent the FR collaborates with.

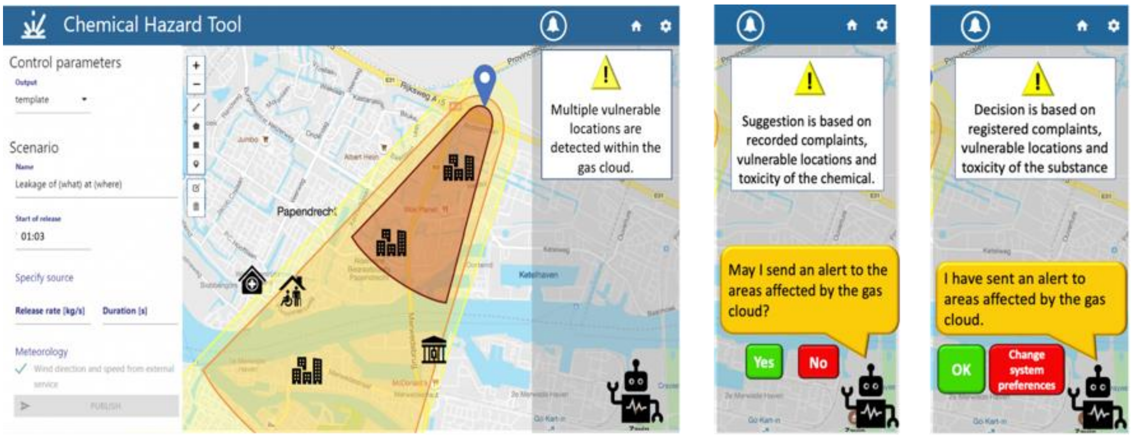


Figure 5. Example of how the different AI agents assist the FR with the decision of whether to send an alert to the public and warn them about the dangerous gas. Left: Informing Agent, Middle: Advising Agent, Right: Deciding

5 Evaluation of Designed Collaboration within the Human-Agent Team

The previous section presented the design of three collaboration styles for the Human-Agent Team (HAT), in form of Team Design Patterns (TDPs), and their implementation into the interface of the Chemical Hazard Tool using Interaction Design Patterns (IDPs). It outlined the first part of work performed to assess the explorative research question of how the collaboration between the FR and Chemical Hazard Tool should be designed. This section describes the second part of work performed, namely the formative evaluation of the designed TDPs (*Informing Agent, Advising Agent* and *Deciding Agent*). The aim of the evaluation was to assess which of the TDPs is most suitable in supporting FRs during the mitigation of incidents. Further, as this work was performed at an early stage of an iterative design process, we included in the formative evaluation the assessment of whether the FRs believed that the designed TDPs lead to a good understanding of the proposed collaboration design and whether they were regarded as reusable for situations other than the once presented during the evaluation.

For the evaluation we compiled an online survey in which FRs were asked to collaborate with the different AI agents in a low fidelity simulation. Below follows a description of participants, scenarios and tasks presented during the low fidelity simulation, procedure during the evaluation, and materials and measures that were employed.

5.1 Participants

The support that the designed AI agents offer is mainly suitable for members of the fire brigade that have a leading function during incident mitigation and who make decisions concerning the action plan. Therefore, we decided to send out the survey to firefighters that met the description above. Participants were recruited through a contact within the fire brigade who distributed an email describing the purpose of the study and inviting the firefighters to participate in the evaluation. The email contained an anonymous link to the survey (Appendix I).

In total 27 participants started the evaluation, of which eight had to be excluded due to not finishing the survey. Hence 19 participants (2 female, 17 male) completed the survey and were suited for analysis. The age of the participants ranged from 34 to 65 years (*M=50.1, SD=10.0*). Years of experience as firefighters ranged from 3 to 40 years (*M=21.3, SD=11.4*). All participants accepted the informed consent (Appendix J). Participants took on average 60 minutes to fill out the questionnaire (*M=60.41*min; *SD=65.43).*

Participants were also asked to indicate the position they were currently holding within the fire brigade. After assessing the dataset, we identified three main positions. These are hazardous materials officers (Hazmat Officers), On-scene Commanders and Safety Policy Officers (NL: beleidsmedewerker). The expertise about hazardous substances varies between these different positions within the fire brigade. Hazmat Officers are experts in dealing with hazardous substances. Their responsibility during incident mitigation is to give advice (usually to the On-scene Commander) regarding the proper handling of the hazardous substance (e.g. what protection gear to wear, how to mitigate the chemical, how to disinfect contaminated people, etc.). They also calculate and predict the hazardous footprint of escaping gas. They instruct measurement teams as to where to do measurements in the affected area to determine toxicity levels in the air and refine the calculated gas cloud distribution. On-scene Commanders lead the mission. They make the main decisions and need to have an overall understanding of what all parties are doing. Safety Policy Officers need to have an understanding of procedures and the work carried out during incident mitigations. They use this knowledge to improve policies, procedures, training and materials[16].

As this is an explorative study and because these three professions vary in their specific knowledge about hazardous substances, we decided to split them into two groups and assess the dataset for possible differences. The first group (11 participants) consists of Hazmat Officers only and the second group (8 participants) combines On-scene Commanders and Safety Policy Officers.

5.2 Scenarios and Decision Points

For the evaluation, two scenarios were created. Both scenarios revolve around incidents in which a toxic gas is released. In one scenario a train carrying chlorine has an accident and is leaking the toxic gas. In the other scenario, an accident at an industrial site causes the escape of the gas sulfur dioxide. Both gasses are harmful when in contact with skin or inhaled. The scenarios are strongly based on the information retrieved from previously performed interviews and the performed document analysis already mentioned in section 4.1. Appendix C presents the chlorine scenario step by step annotated in the format of a use case.

Throughout both scenarios the firefighters encounter tasks in which they have to make decisions. These decisions are supported by the different AI agents according to the designed

[16] This information was retrieved from interviews that were held and from the HAZMAT officer handbook: Brandweer Kennis Net. Brandweer Flevoland. (2014). Accessed: July 18, 2021 [online]. Available: https://www.ifv.nl/kennisplein/cbrn/publicaties/handboek-meetplanleider

TDPs. The decision points are the same for both scenarios and were already mentioned as tasks in section 4.1. Firefighters had to decide:

- which route to take to the incident location;
- whether to alert the public about the escaping gas and recommend closing windows and staying inside;
- where in the affected area measurements should be made to determine toxicity levels and reduce the uncertainty of the calculated gas cloud distribution. This decision had to be made twice;
- whether to warn firefighters that are located within the calculated danger zone.

Appendix K shows how the different AI agents assisted the firefighters at each of the decision points in both scenarios. Above mentioned decision points were chosen because on the one hand, firefighters have to make these decisions during real life missions. And on the other hand, they can be classified as joint activities in which the Chemical Hazard Tool is able to offer assistance.

The first decision point addresses the intention of firefighters to choose a safe route to the incident location that avoids driving directly through a potentially toxic gas cloud.

The second decision is related to the duty of firefighters to keep the public safe and therefore having to estimate the potential threat of a gas that is spreading in the air. Depending on different meteorological data (e.g., wind direction and wind speed, temperature, rainy or sunny conditions, etc.), the toxicity of the gas and identified vulnerable locations (e.g. densely populated areas, hospitals, nursing homes, etc.), firefighters decide whether to take actions like warning the public about the threat and giving instructions on how to stay safe (e.g., closing windows, staying inside, etc.). The Netherlands has a system, called NL-Alert[17] which sends messages directly to the smart phones of everyone who is located within a specific area. Firefighters can make use of that system to alert citizens.

The third decision deals with the inherent uncertainty of calculating the trajectory of a gas cloud, e.g. due to uncertainty about the exact release rate, atmospheric stability, wind speed and wind direction or also sometimes about which exact gas or gases are escaping. To get a more accurate estimation, measurement teams are sent out to specific points in the affected area to measure the concentration of the gas in the air. A designated firefighter, usually the

[17] https://crisis.nl/nl-alert

Hazmat Officer, assigns the measurement positions to the teams (Brandweer Flevoland, 2014). Within the scenario the firefighters are asked twice to allocate measurement position to measurement teams. The first time before any measurements in the field had been performed and the second time to reassign the measurement team after performing the first measurement assignment. This usually also happens during a real-life mission, i.e., once the measurement team has done the assigned measures, it is usually reassigned to a new position or sometimes put on standby.

The fourth decision is related to the fact that firefighters do not only intend to keep the public safe but also themselves and their team members. When firefighters are sent out in the field some areas are more dangerous than others, e.g. due to unexpected gas concentrations and lack of protective gear. Therefore, if detected that some team members are within or in the vicinity of a dangerous area, it can be decided to warn them.

5.3 Procedure

As stated above the link to the online survey was distributed via email. Once participants opened the link the survey started. Figure 6 illustrates in a flow chart the steps participants went through when filling out the survey. First, participants were informed about the aim of the study and were asked to read and accept the informed consent (Appendix J). This was followed by asking the participants to indicate demographic data.

Then the collaboration simulation started which was split into two blocks, as can be seen in Figure 6. In the first block participants went through two rounds of collaboration simulation. In the first round they either collaborated with the *Informing Agent* or with the *Advising Agent* to solve either the Chlorine Incident or the Sulfur Dioxide Incident. In counterbalanced order participants collaborated once with each of the AI agents while solving the Chlorin Incident in one case and the Sulfur Dioxide Incident in the other case. Before the scenario started, participants were informed about the capabilities of the Chemical Hazard Tool and about the Team Design Pattern (TDP) they were about to collaborate in. This included showing a depiction of the TDP and providing a description of the roles and responsibilities within the TDP (see Appendix L & Appendix M). Once the scenario of the first round was solved the firefighters were asked to rate the perceived collaboration with the according AI agent, the scenario itself and their situation awareness. In the second-round participants collaborated with the other AI agent and solved the incident they had not yet encountered. For example, if in the first round a participant collaborated with the *Informing Agent* to solve the Chlorine Incident, then in the second-round s/he would collaborate with the *Advising Agent* to solve the Sulfur Dioxide Incident. At the end of the first block participants were asked to indicate which AI agent they preferred to be assisted by during the different decision points.

Then Block 2 started. There participants were introduced to the *Deciding Agent* (Appendix N). This time they did not play through a whole new scenario. We presented them three decisions (alerting the public, allocating measurement teams, warning team members) from the Chlorine Incident. Though, this time participants were shown how the collaboration with the *Deciding Agent* would have looked like. At each of the presented decisions points participants were asked whether they would accept that the *Deciding Agent* continues making decisions on its own for similar decisions or if they would prefer to change the TDP to either the *Informing Agent* or the *Advising Agent* (an example is shown in Appendix H).

After the second block, participants were asked to indicate final comments about each of the AI agents. Then the survey was completed. Participants were thanked for their participation and were provided with contact details in case they wanted to request more information.

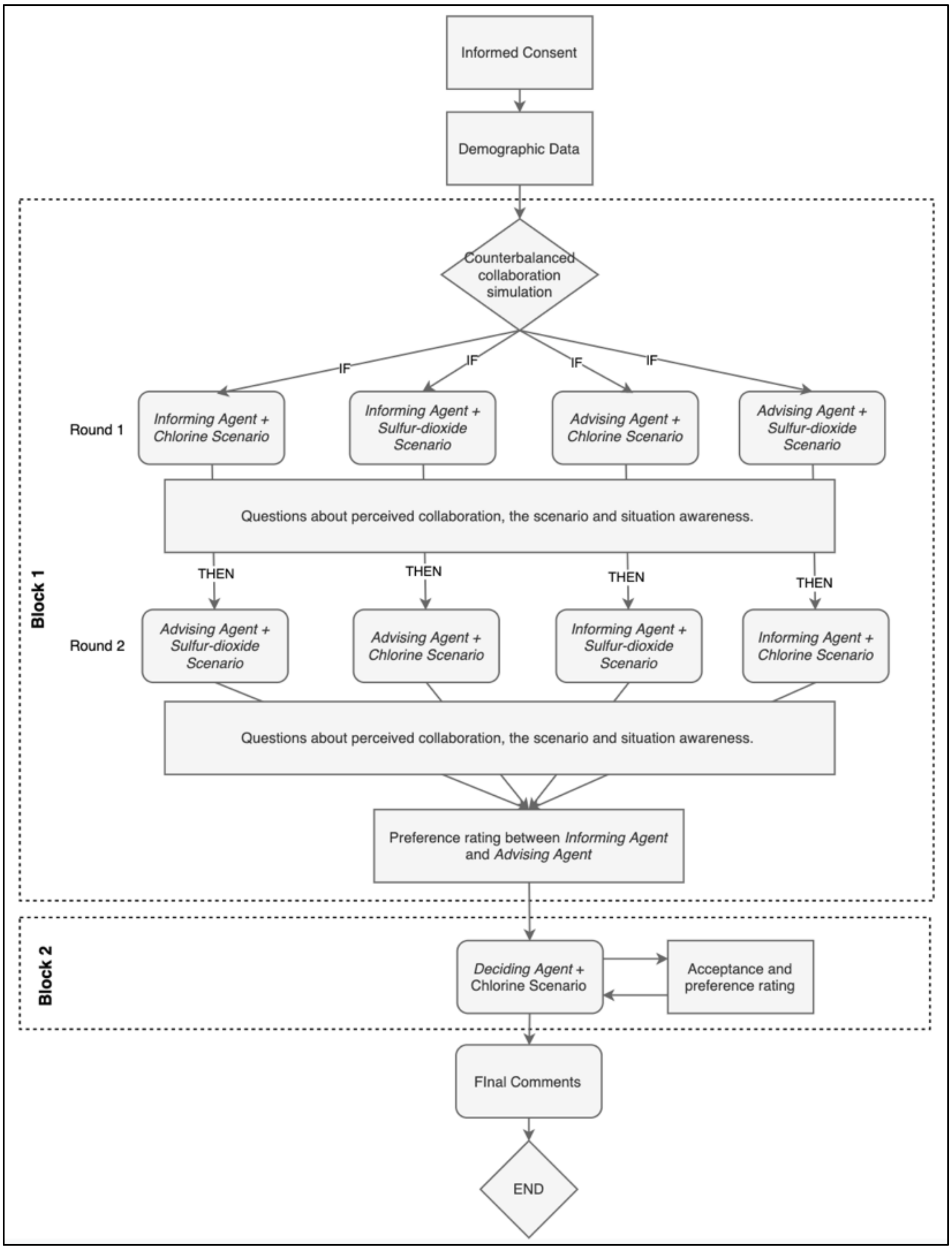


Figure 6. Survey Flow. The figure shows the steps that participants go through when filling out the survey. First, they fill out the informed consent, followed by indicating their demographic data. Then the first block of collaboration simulation begins. The participants collaborate in one of the four combinations with either the Informing Agent or the Advising Agent and solving either the chlorine incident or the sulfur dioxide incident. Once the incident is solved, they are asked questions about the perceived collaboration with the AI agent, the scenario they had to deal with and their situation awareness. Then they go through a second simulation but this time working together with the other AI agent and solving the incident they have not encountered yet. Afterwards they are asked again questions about the perceived collaboration, the scenario and situation awareness but this time according to the other constellation they experienced. At the end of this first block, participants are asked to indicate whether they preferred to work with the Informing Agent or the Advising Agent. Then the second block starts, there the participants are introduced to the Deciding Agent but instead of solving a new scenario they revisit the chlorine incident and are shown for three decision points how the collaboration with the Deciding Agent would have looked like. At each of these deciding points the participants are asked whether they would accept the Deciding Agent or if they would prefer to collaborate with either the Informing Agent or Advising Agent.

5.4 Material and Data Analysis

The platform www.qualtrics.com was utilized to create the online survey and collect data. The survey was made available in English and Dutch. The Dutch version was chosen by all participants that completed the survey.

The survey flow depicted in Figure 6 shows that the survey was split into two blocks. In the first block the *Informing Agent* and the *Advising* Agent were presented and in the second block the *Deciding Agent* was introduced. Participants were asked to solve two scenarios in the first block. One scenario included the escape of the toxic gas sulfur dioxide at an industrial plant. In the other scenario a train carrying the toxic chemical chlorine crashed and was leaking the gas. In the second block the chlorine scenario was revisited, this time showing how the collaboration with the *Deciding Agent* would have looked like for specific decision points within the scenario. Throughout the survey different measures were implemented to investigate how we should design the collaboration between the Chemical Hazard Tool and the first responder.

As presented in the introduction we formulated the following research question:

"How should the collaboration between first responders and the Chemical Hazard Tool be designed? Which of the created Team Design Patterns is suitable to support first responders and assures adequate situation awareness?".

Additionally, we formulated the following research sub-question:

"Do the created TDPs lead to a good understanding of the proposed collaboration designs among the first responders and do the first responders regard the TDPs as reusable?"

To answer the research sub-question of whether TDPs indeed meet the claim of being a good way to communicate design choices and create a common ground for discussing these choices, questions were presented that query the understandability and reusability of the displayed collaboration. Participants were asked to rate the statements on a five-point Likert scale from "strongly disagree" to "strongly agree". The questions were based on the work performed by Van Stijn (2020. The language of the questions was adapted from addressing researchers to addressing first responders and disaster mitigation settings. The questions are presented in Appendix D. Descriptive statistics were used to see on average how much participants agreed with the statements. Inferential statistics, to be exact the Wilcoxon Signed-Rank Test, was used to assess possible differences between the TDPs.

In general, whenever the data set is assessed with inferential statistical tests, we use non-parametric tests, because of the relatively small sample size (N=19) and due to the scales of measure being usually either ordinal (e.g. Likert scales) or nominal.

To answer the main research question, the firefighters were presented with incident scenarios which they solved in a low fidelity simulation. As this is a formative evaluation using an explorative approach, we decided to assess the legitimacy of the presented scenarios. The firefighters were asked to indicate on a five-point Likert scale how often they had experienced similar incidents in the past (from “never” to “always”) and whether they rated the scenarios as realistic (from “strongly disagree” to “strongly agree”). The items are presented in Appendix E. They were implemented into the survey as a control measure to make sure that the TDPs were tested within scenarios that were considered realistic. Further, the items served as an additional check to an interview held with a firefighter during the design of the evaluation in which the authenticity of the scenarios was discussed. Descriptive statistic was used to describe the general impression of the firefighters. Inferential statistics, in this case the Wilcoxon Signed-Rank Test, was used to assess differences between the presented scenarios.

As the main research question queries which TDP assures adequate situation awareness, two measures of situation awareness were employed within the survey. We used the style of the explicit measure Situation Awareness Global Assessment Technique (SAGAT) developed by Endsley (1988). On two occasions the scenario was interrupted, the screen blanked and questions about the current situation were asked. The accuracy of the answers was assessed by comparing given answers to the actual situation. Percentages of correct answers are calculated to illustrate the correctness rate. Additionally, subjective SA was assessed by deriving and adjusting questions from the Mission Awareness Rating Scale (MARS, Matthews & Beal, 2002). In total four items (one workload and three content items) were adapted from the MARS questionnaire. Participants were asked to rate the items on a five-point Likert scale ranging from “strongly disagree” to “strongly agree” (Appendix F). Descriptive statistics was used to describe the average situation awareness and inferential statistics (Wilcoxon Signed-Rank Test) was used to test for differences between the Team Design Patterns.

To further answer the main research question and assess whether the AI agents are suitable to support the firefighters during mission mitigation, we asked them to rate the helpfulness of the *Informing Agent* and the *Advising Agent* on a five-point Likert scale from “strongly disagree” to “strongly agree” (see Appendix D). Again, descriptive statistics was used to show the general opinion and the Wilcoxon Signed-Rank Test was used to assess whether differences could be found between the two AI agents.

With the same purpose of assessing how and when the AI agents are offering suitable support to the firefighters, we recorded whether the participants would follow the advice given by the *Advising Agent.* At each of the decision points (route to the incident location; alerting the public; assigning measurement positions; warning team members located in a danger zone) the *Advising Agent* would provide a suggestion and the participants had the option to either accept the suggested solution or decline it and apply their own solution. The rate with which suggestions were followed for each of the decision points was assessed by calculating percentages. Additionally, differences between the two identified groups of participants (Hazmat Officers vs. Commanders and Policy Officers) were assessed. As the two groups usually differ in expertise and responsibilities during real incident mitigation, we decided to assess whether a difference could be found in the willingness to follow the advice of the AI agent. Differences between the two groups were assessed using the Chi-Square Test.

To further gather data and answer the main research question about which AI agent would be most suitable to assist the firefighters, we asked them to indicate which collaboration style they preferred after having completed Block 1 and therefore having collaborated with the *Informing Agent* and the *Advising Agent*. For each of the four decision points firefighters had the option to choose between the options: "*Informing Agent", "Advising Agent"*, "Both are alright" and "Neither". The items are presented in Appendix G. The participants indicated their preference for each of the four decision points. The frequency with which the options were chosen was assessed through percentages and whether the difference between the options was significant was assessed using the Chi-Square Test.

In the second block it was assessed whether the firefighters accepted the *Deciding Agent* as a collaboration partner or if they preferred to rather collaborate with the *Informing Agent* or *Advising Agent.* This measure was also put in place to find an answer to the main research question and assess which of the three AI agents was regarded as suitable to assist the firefighters. For this purpose, participants revisited three moments in the chlorine incident where they had to make incident relevant decisions. This time the participants were shown how the collaboration with the *Deciding Agent* would have looked like. They were given the choice to accept collaborating with the independently acting *Deciding Agent* and allow it to continue making decisions for similar situations or to change the collaboration to either TDP 1 (*Informing Agent*) or TDP 2 (*Advising Agent*). An example is given in Appendix H. For each task it was assessed which AI agent was preferred, this was done to gauge whether a relation between task at hand and preferred AI agent could be identified. Further it was assessed whether differences could be identified depending on the group (Hazmat Officers vs. Commanders and Policy Officers) the participant belonged to. Differences between the frequency with which the AI agents were chosen was assessed using the Chi-Square Test. To further gain insight into the reasoning of the firefighters when filling out the survey and to gather further opinions and requirements for the AI agents, we asked the firefighters to provide comments, explanations and suggestions throughout the whole survey.

6 Results

This section presents the results obtained from the performed survey of which the aim was to assess which AI agent would be best suitable to support FRs during their mission. First the results from Block 1 are presented, then from Block 2 and at the end, an overview of the comments provided by the firefighters will be given. Statistical tests were performed using nonparametric tests due to the relatively small sample size (N=19) and due to the scales of measure being either ordinal (e.g. Likert scales) or nominal. Statistical significance is reported at a threshold of $\alpha = 0.05$. Only descriptive statistics are reported when statistical significance was not given but tendencies could still be identified.

The results obtained from Block 1 are understandability and reusability measures of the presented Team Design Pattern; judgments about how realistic the presented scenarios were; explicit and subjective situation awareness ratings; perceived helpfulness of the *Informing Agent* and the *Advising Agent;* number of times the advice of the *Advising Agent* was followed and indicated preference between collaborating with the *Informing Agent* and the *Advising Agent.*

Results obtained from Block 2 provide information about acceptance towards the *Deciding Agent* and preferences between the three AI Agents.

Below section 6.1.1 Understandability and section 6.1.2 Reusability present the results related to the research sub-question: "Do the created TDPs lead to a good understanding of the proposed collaboration designs among the first responders and do the first responders regard the TDPs as reusable?".

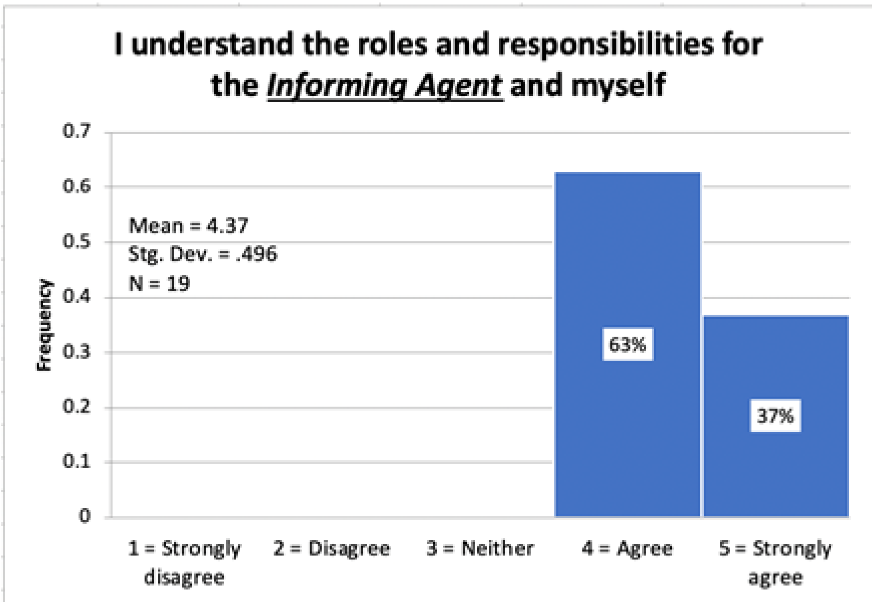
Section 6.2 presents the results related to the main research question: "How should the collaboration between first responders and the Chemical Hazard Tool be designed? Which of the designed Team Design Patterns is suitable to support first responders and assures adequate situation awareness?".

6.1 Results related to the Research Sub-Question

6.1.1 Understandability

The distribution of understandability scores for TDP 1 (*Informing* Agent) and TDP 2 (*Advising Agent)* are shown in Figure 7. Participants rated on a five-point Likert scale from 1 = "Strongly disagree" to 5 = "Strongly agree" whether they understood the roles and responsibilities that were assigned to them and the AI agent. The explanation of allocate roles and responsibilities for the *Informing Agent* and the *Advising Agent* was done in the same style for both. Participants were shown the graphical representation presented in Table 2 and Table 3 and given a verbal description. Appendix L and Appendix M show the explanation participants received during the survey. The results show that on average participants understood well how the roles and responsibilities were allocated within both TDPs (*Informing Agent:* M =4.37, *SD*= .496 and *Advising Agent*: *M* = 4.26, *SD*=.562). Further, employing the Wilcoxon Signed-Rank Test showed that the roles and responsibilities within the HAT were equally well understood for both AI agents as no significant difference was found (Z = -1.41; p = .157).

A.



B.

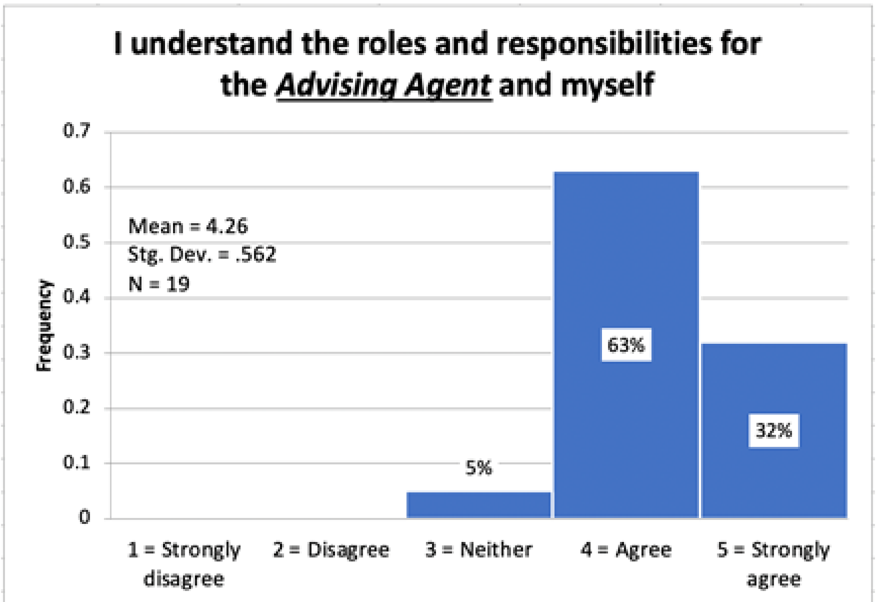


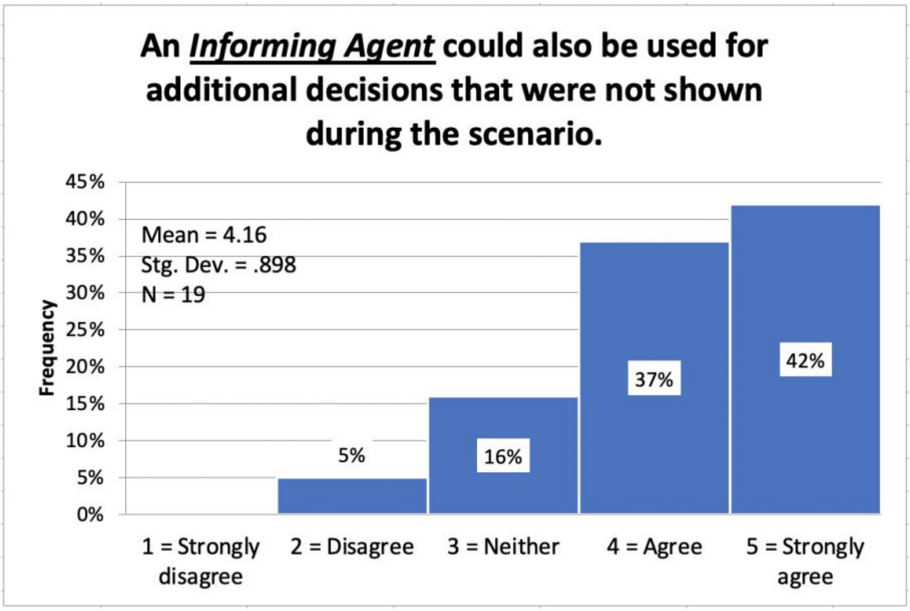
Figure 7. Rated understandability of presented roles and responsibilities within the Human-Agent Team using Team Design Patterns. Ratings were given on a five-point Likert Scale from 1 = "Strongly disagree" to 5 = "Strongly agree". On

the top (A): Accumulated responses for understandability of TDP 1 (Informing Agent). On the bottom (B): Accumulated responses for understandability of TDP2 (Advising Agent).

6.1.2 Reusability

Histograms showing the distribution of reusability scores for the collaboration described in TDP 1 (*Informing Agent*) and TDP 2 (*Advising Agent*) are presented in Figure 8. Addressing the research sub-questions, participants were asked to rate on a five-point Likert scale (from 1 = "Strongly disagree" to 5 = "Strongly agree") whether they agreed that the assigned roles and responsibilities presented in the TDPs could also be applied for other tasks and decisions that were not encountered during the scenario. Participants agreed on average with the statement above (*Informing Agent*: M=4.16, SD=.898; *Advising Agent*: M=4, SD=.943). Again, after preforming the Wilcoxon Signed-Rank Test, the results showed no significant difference (Z = -1.34; p = .180) between the two AI agents, meaning that participants judged on average the collaboration with both AI agents as equally transferable to tasks and decisions that were not presented during the scenarios.

A.



B.

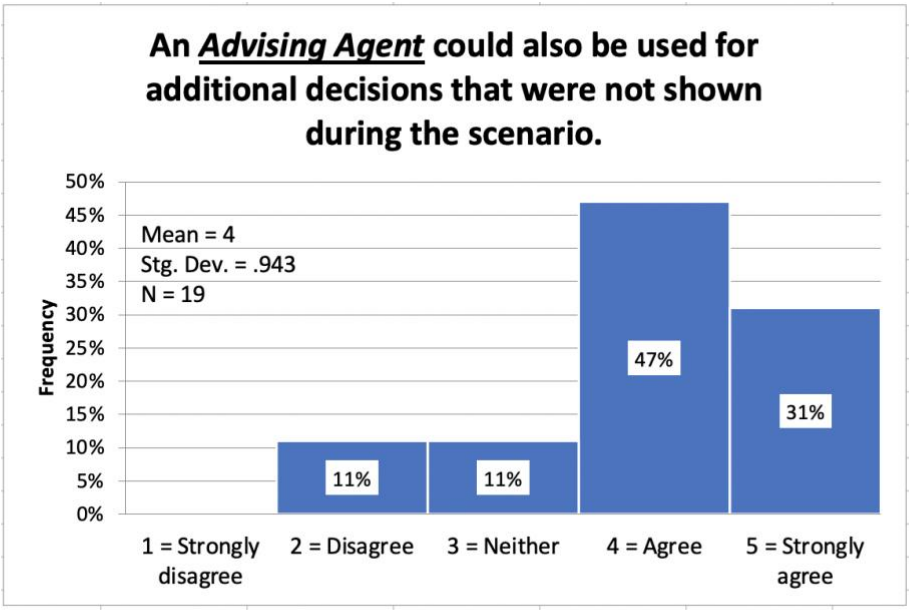


Figure 8. Rated reusability of TDP 1 (Informing Agent) on the top (A) and TDP 2 (Advising Agent) on the bottom (B). Ratings were given on a five-point Likert Scale from 1 = "Strongly disagree" to 5 = "Strongly agree".

6.2 Results Related to the Main Research Question

6.2.1 Scenarios

Participants solved two scenarios during the survey, one was dealing with a train accident which led to a leakage of chlorine and the second was dealing with an incident at an industrial side that led to the escape of sulfur dioxide. Participants were asked to rate how often they had experienced similar situations in the past on a five-point Likert scale (from 1 = "Never" to 'Always' = 5). For both scenarios the frequency of having experienced similar incidents in the past was low. When comparing the results for both scenarios using the Wilcoxon Signed-Rank Test no significant difference was found (Chlorin Scenario: M = 1.89, SD = .937; Sulfur Dioxide Scenario: M = 2.21, SD = .713; Z = -1.732; p = .083).

Even though participants were unlikely to have encountered similar incidents in the past, when asked to rate on a five-point Likert scale (from 1 = "Strongly disagree" to 5 = "Strongly agree") whether they regarded the presented scenarios as realistic, they indicated on average to agree with the statement. They viewed both scenarios as equally realistic as there was no significant difference found using the Wilcoxon Signed-Rank Test (Chlorine Scenario: M = 3.84, SD = .898; Sulfur Dioxide Scenario: M = 4, SD = 1.04; Z = -.905; p = .366).

6.2.2 Situation Awareness

Situation Awareness (SA) was measured explicitly and subjectively. For the explicit SA assessment, the screen was blackened during the scenario and the firefighters were asked incident-relevant questions. The answers were then compared to the actual situation and a SA score was calculated. Participants achieved overall high SA scores, responding correctly to 88% of the question when collaborating with the *Informing Agent* and responding correctly to 87% of the questions when collaborating with the *Advising Agent.* No significant difference was identified.

Subjective SA was assessed based on the Mission Awareness Rating Scale (MARS, Matthews & Beal, 2002), from which four items were adapted. The four presented items were "It was easy to predict what was about to occur", "It was easy to make mission relevant decisions", "It was easy to identify mission-critical information during the scenario" and "It was easy to identify mission-critical information during the scenario". Participants rated the items on a five-point Likert scale (from 1 = "Strongly disagree" to 5 = "Strongly agree") at two occasions during the survey, once after collaborating with the *Informing Agent* and once after collaborating with the *Advising Agent*. The results for the four individual items are presented in Figure 9. The combined items represent the overall subjective SA score, which was on average acceptable for both collaboration styles (*Informing Agent*: M = 4.04, SD = 0.74; *Advising Agent*: M = 3.97, SD = 0.86). When comparing this overall subjective SA score using the Wilcoxon signed Rank test no significant difference between the two Team Design Patterns was found (Z = -.713; p = .476).

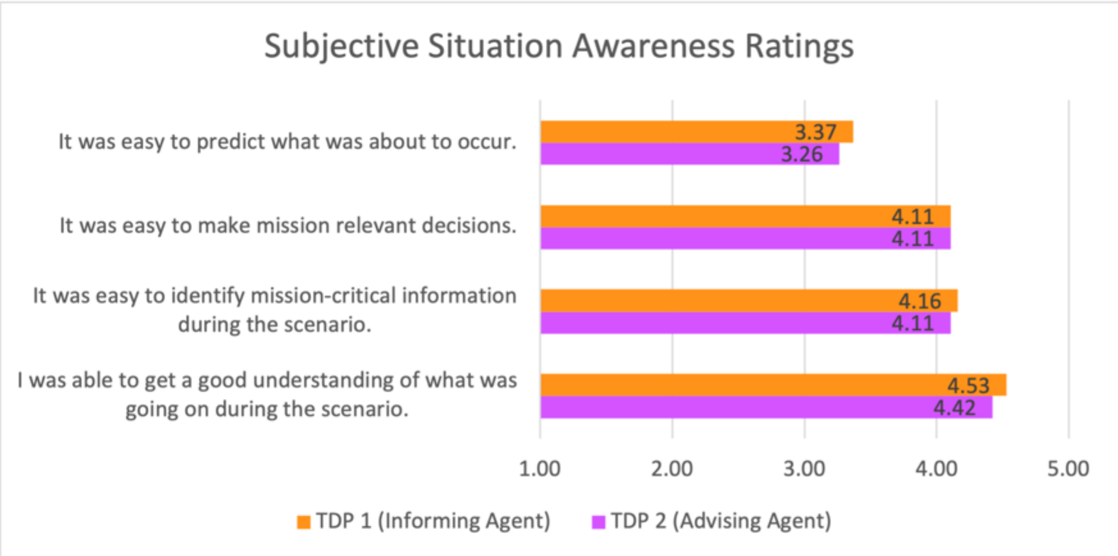


Figure 9. Subjective SA ratings. Participants rated four items on a five-point Likert scale from 1 = "Strongly disagree" to 5= "Strongly agree". The chart shows the average ratings of the subjective SA items indicated after collaborating with the Informing Agent and after collaborating with the Advising Agent.

6.2.3 Perceived helpfulness of the Agents

How helpful participants rated the support of the *Informing* Agent and the *Advising Agent* when solving the scenario is shown in Figure 10. The statement that the AI agents were helpful was rated on a five-point Likert scale from 1 = "Strongly disagree" to 5 = "Strongly agree". Overall, it was agreed that the support of both AI agents was helpful (*Informing Agent*: *M* = 4.26, *SD* = 0.73; *Advising Agent*: *M* = 4.16, *SD* = 0.90) and no significant difference could be found using the Wilcoxon Signed-Rank Test (*Z* = -1.0, *p* = .317). Looking at Figure 10, it can be seen that one participant disagreed in the case of the *Informing Agent* and two participants disagreed in case of the *Advising Agent*. One of these participants remarked that to offer better support the AI agents should incorporate the weather forecasts much stronger in their support.

A.

B.

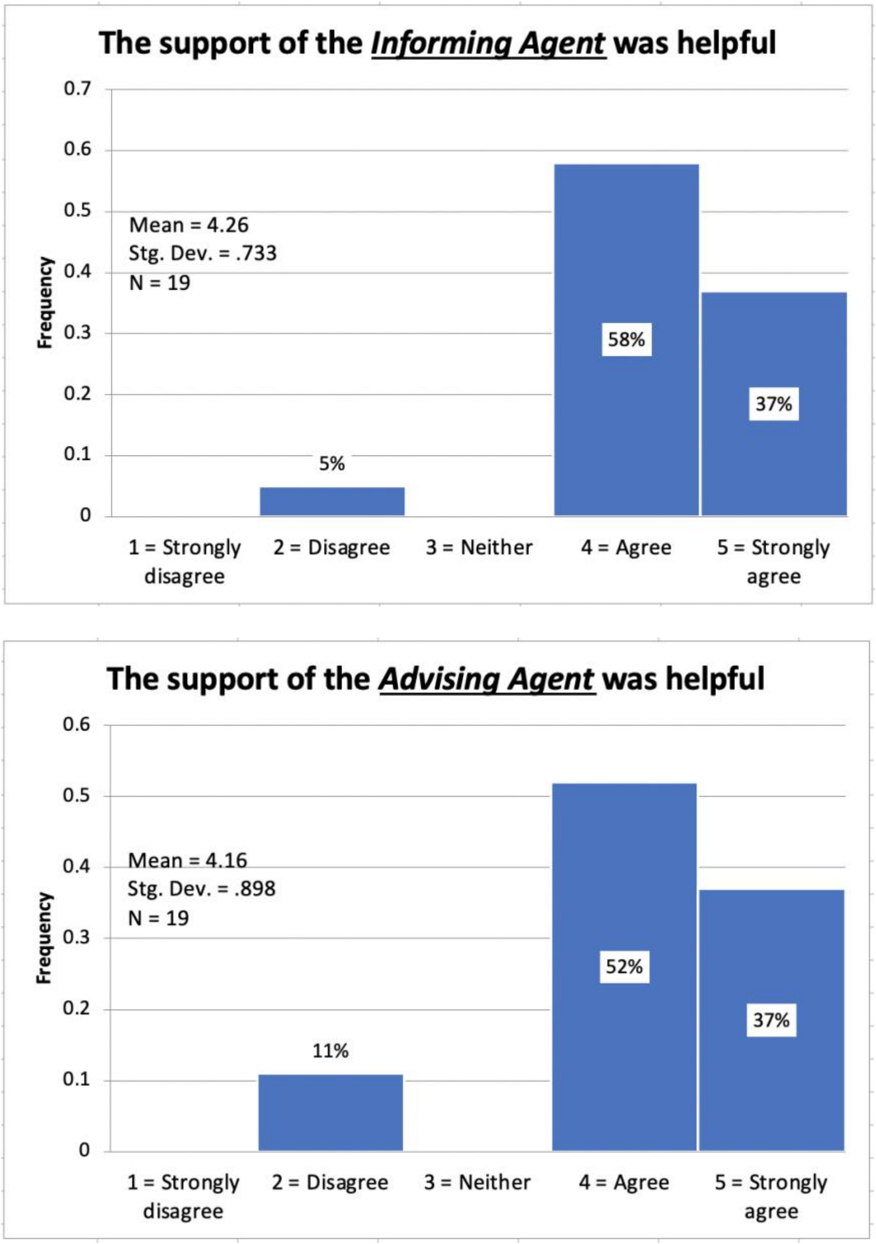


Figure 10. Rated helpfulness of the support offered by the Informing Agent (image A) and the Advising Agent (image B). The statements were rated on a five-point Likert scale from 1 = "Strongly disagree" to 5 = "Strongly agree".

6.2.4 Following the Advice of the *Advising Agent*

During the survey participants had to make a number of decisions. These included choosing a safe route to the incident location, deciding whether to alert the public, deciding where to send measurement teams (2x) and deciding whether to warn team members that are located within a danger zone (see section 5.2). When collaborating within TDP 2, the *Advising Agent* gave recommendations for all these decisions. Figure 11 shows how often the given advice was accepted by the participants. The results are shown separately for the two groups we identified within the dataset (see section 5.1). Group 1 consists of Hazmat Officers and Group 2 combines On-Scene Commanders and Policy Officers.

In total, the recommendations given by the *Advising Agent* were followed 76% of the time. Differences were found depending on the decision the participants had to make and the group they belong to. All participants accepted the suggested route to the incident location. In contrast, the recommended measurement positions were followed the least often with an acceptance rate of 63%. Measurement positions had to be indicated on two occasions, the results on the acceptance of the advised measurement positions were exactly the same for both occasions and are therefore reported together. Participants that did not accept the suggested measurement positions commented that they wanted more information about the measurement strategy that the *Advising Agent* was basing the suggestions on, and some wanted to be able to adjust the measurement strategy to match their own preferences.

Assessing the difference between the two groups using the Chi-Square Test showed a significant difference (χ^2 (1, N = 95) = 5.164, p = .023) between the total acceptance rate of Group 1 compared to Group 2. Group 1 (Hazmat Officers) accepted 67% of the provided advices which is less than the acceptance rate of Group 2 (Commanders & Policy Officers), which accepted 87.5% of the provided advices. For each of the individual decisions (route to location, alert public, measurement positions, warn team members) the small sample size of 19 answers contributed to not finding significant differences, therefore only tendencies are described. The difference was most prominent for the decision whether to accept the advice to alert the public. In Group 1 the advice was followed by 54% of the Hazmat Officers and in Group 2 the advice was followed by 87% of the Commanders and Policy Officers. Participants that did not follow the advice to alert the public remarked that they wanted more insight about what information the *Advising Agent* based its advice on and which exact area the alert would be sent to.

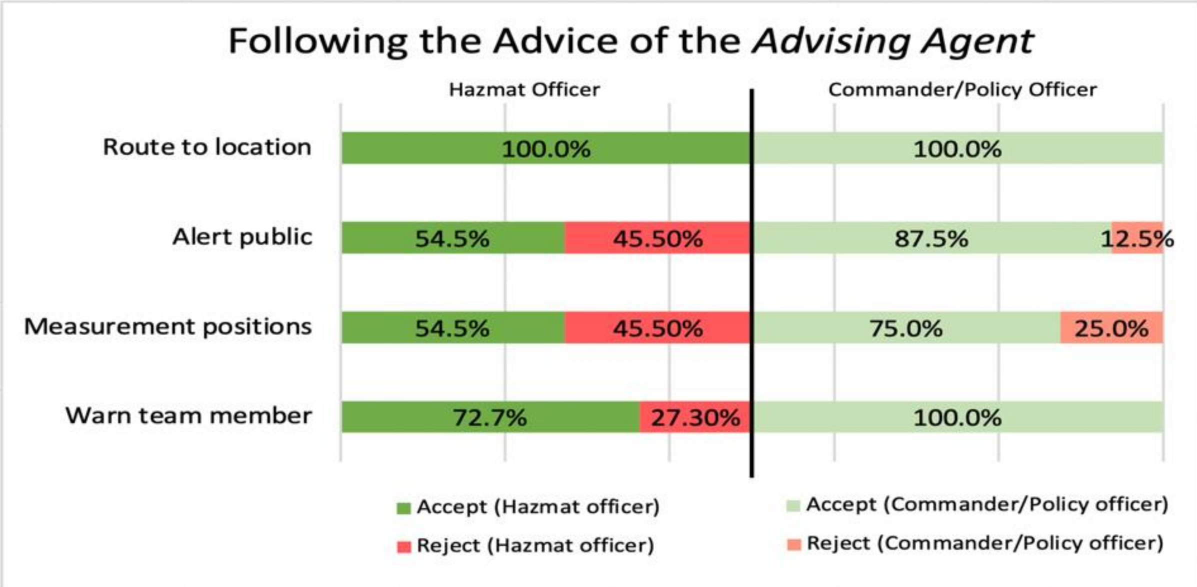


Figure 11. Indication of how often the recommendations of the Advising Agent *was followed. Four decision points are listed (measurement positions were indicated twice during the survey but are shown here once because the results were the same). Green indicates the percentage for advice followed and red indicated the percentage for advice rejected. On the left the results for Group 1 (Hazmat Officers) are shown. On the right, the results for Group 2 (Commanders and Policy Officers) are shown.*

6.2.5 Preference Ratings after Completing Block 1 of the Survey

At the end of the first block, participants were asked to indicate which collaboration style they preferred. For each of the decision points (i.e., route to the incident location, alerting the public, allocating measurement teams and warning team members) they were given the option to vote for either the "*Informing Agent*", the "*Advising Agent*", "Both are alright" (meaning that they have no preference) or "Neither" (meaning that they would prefer to work without the assistance of an AI agent). The results of this preference rating are shown in Figure 12.

Overall, the results show that participants preferred to be rather assisted by one of the AI agents than not to receive assistance, as only 5% of the overall preference votes indicated the option "Neither". "Both are alright" received 37% of the total votes, "*Advising Agent*" received 36% of the total votes and "*Informing Agent*" received 22% of the total votes. Using the Chi-Square test, we found that the difference between not wanting to be assisted by an AI agent compared to wanting to be assisted is significant (χ^2 (3, N = 19) = 19.684, p = .000).

When examining the preference ratings for each decision point (see Figure 11), no uniform tendency for preferences between "*Informing* Agent", "*Advising Agent*", "Both are alright" and "Neither" can be identified. This means that no statement can be made about which TDP should be matched with which task.

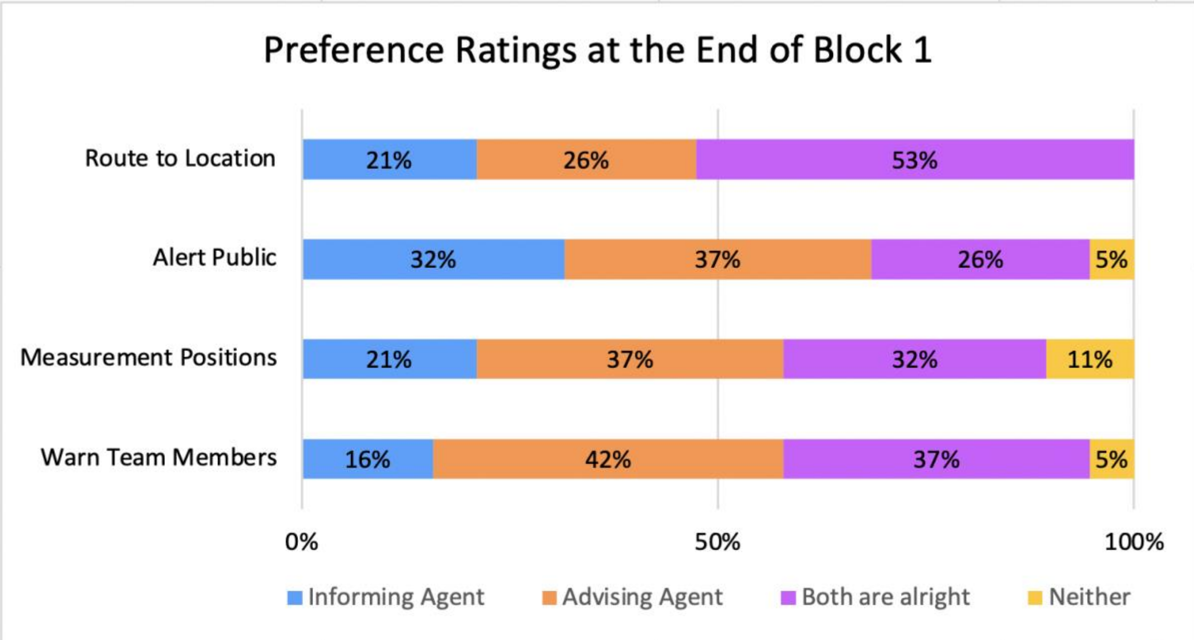


Figure 12. Preference ratings after collaborating with the Informing Agent *and the* Advising Agent*. Participants were asked to indicate their preference for each of the four decision points (deciding on a route to the incident location, deciding whether to alert the public, deciding on measurement positions and whether to warn team members located in a danger zone). Participants could, choose between the "Informing Agent", the "Advising Agent", "Both are alright" and "Neither". Percentages indicated how many participants voted for the option to which the respective color belongs to.*

We also looked at the preference rating for each of the participants to assess whether they tended to vote for the same AI agent for all tasks or whether depending on the task they preferred different AI agents to collaborate with. Figure 13 shows the results for preferences indicated by each participant. The results show that participants predominantly vary their preferences and do not stick to the same collaboration style for all tasks. This can be seen as only two participants indicated to prefer to collaborate with the same AI agent on all presented tasks. In additional three participants indicated "Both are alright" for all tasks, which does not give insight to whether they would have liked to collaborate with only one AI agent or with varying AI agents.

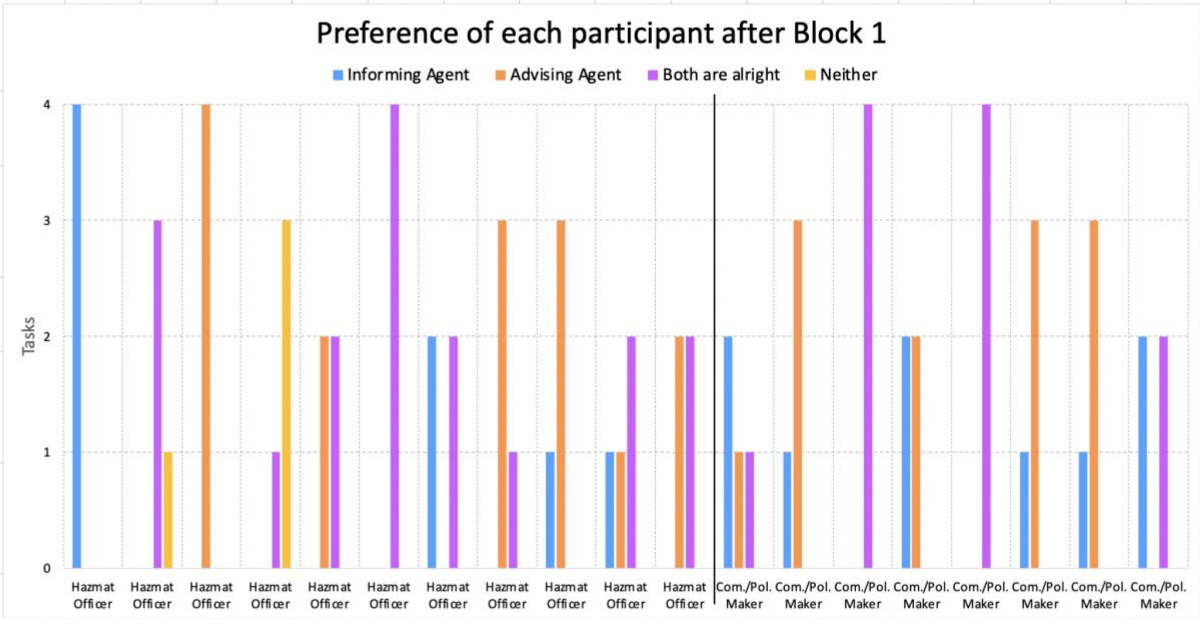


Figure 13. Indicated preferences of each participant after Block 1. The graph shows how many times each participant, voted for one of the following options: "Informing Agent", "Advising Agent", "Both are alright" or "Neither". It further indicates whether a participant belongs to Group 1 (Hazmat Officers) on the left or Group 2 (On-scene Commanders and Policy Officers on the right). Participants mostly varied in their preferences across different decision points and did not stick to one option.

6.2.6 Comparison between the Deciding Agent, the Advising Agent and the Informing Agent

During Block 2 participants were shown how the collaboration with the *Deciding Agent* (in TDP 3) would look like. Three decisions from the Chlorine Scenario were revisited for that purpose. The shown decisions were deciding whether to alert the public, deciding where to send measurement teams and deciding whether to warn team members located within a calculated danger zone. Participants were asked to indicate whether they would give the *Deciding Agent* the permission to continue deciding on its own for these types of decisions or whether they preferred to switch to collaborating with the *Informing Agent* or the *Advising Agent.* In contrast to the previous section 6.2.5, here all three AI agents are compared.

Figure 14 shows the response rates for each of the three tasks. A chi-square test was used to assess which of the AI agents was regarded most suitable for each of the tasks at hand. The tasks are assessed separately to identify whether depending on the task different AI agents were chosen.

For the task warning team members located in a calculated danger zone, a significant difference was found (χ^2 (2, N = 19) = 14.000, p = .001) between the frequency with which the three AI agents were chosen. Most participants (74%) agreed to give the *Deciding Agent* permission to autonomously warn team members when it registered that the team members entered a danger zone.

Concerning the task *Alerting the Public*, again a significant difference was found for the frequency with which the AI agents were chosen by the participants, χ^2 (2, N = 19) = 14.632, p = .001. Most participants (74%) choose to collaborate with the *Advising Agent* when deciding whether to alert the public.

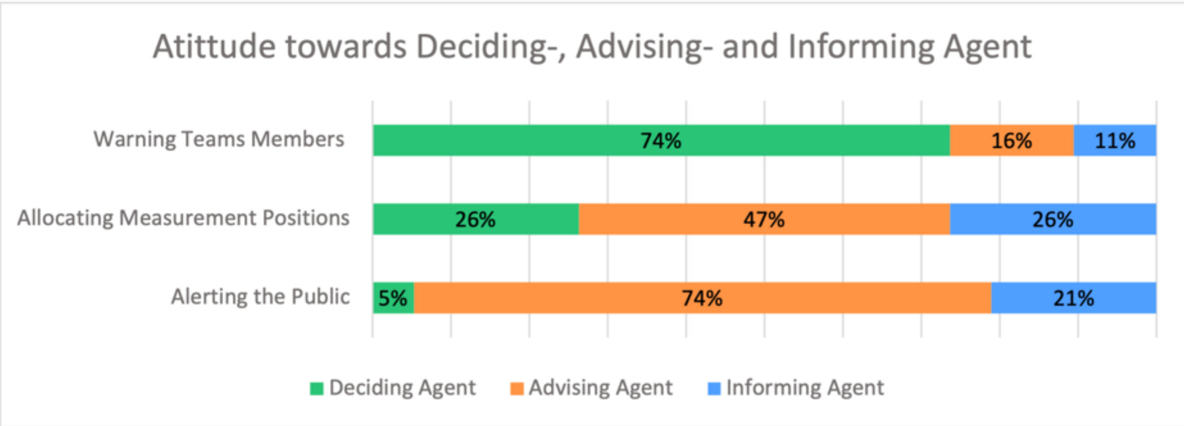


Figure 14. Attitude towards collaborating with the Deciding Agent for three decision points. The graph indicates for each of the three decisions (i.e., deciding whether to warn team members located in danger zones, deciding on measurement positions and deciding whether to alert the public) how many participants accepted to collaborate with the Deciding Agent and how many chose to switch to the Advising Agent or the Informing Agent.

For the task allocating measurement positions to the measurement teams, no significant difference (χ^2 (2, N = 19) = 1.684, p = .431) was found regarding the preference between the three AI agents. Nevertheless, Figure 15 shows differences in tendencies that were observed between the two groups within the dataset (Group 1 = Hazmat Officers, Group 2 = On-Scene Commanders and Policy Officers).

Group 1 (Hazmat Officers) tended to accept the *Deciding Agent* less often than Group 2 (Commanders and Policy Officers). Also, Group 1 tended to choose the *Informing Agent*, which offers the least active support in assigning measurement teams, more often than Group 2. Both groups tended to choose the *Advising Agent* most often.

Like for the preference ratings after completing Block 1, we also assessed for the ratings during Block 2 if participants tended to prefer collaborating with the same AI agent for each of the decisions presented or whether they chose different collaboration partners depending on the decision at hand. Again, the results show that there was no uniform approach, some participants chose the same AI agent for all three decision points and other varied depending on the decision.

Participants were asked to provide comments whenever they did not accept that the *Deciding Agent* acts autonomously. We identified recurring themes in the provided comments. One theme was wanting to stay in control of decisions that were made. Another theme was requesting more insight about what information the *Deciding Agent* based its decisions on. Further, some participants doubted that the *Deciding Agent* would have the capability to adapt accordingly to the highly variable environments the firefighters work in.

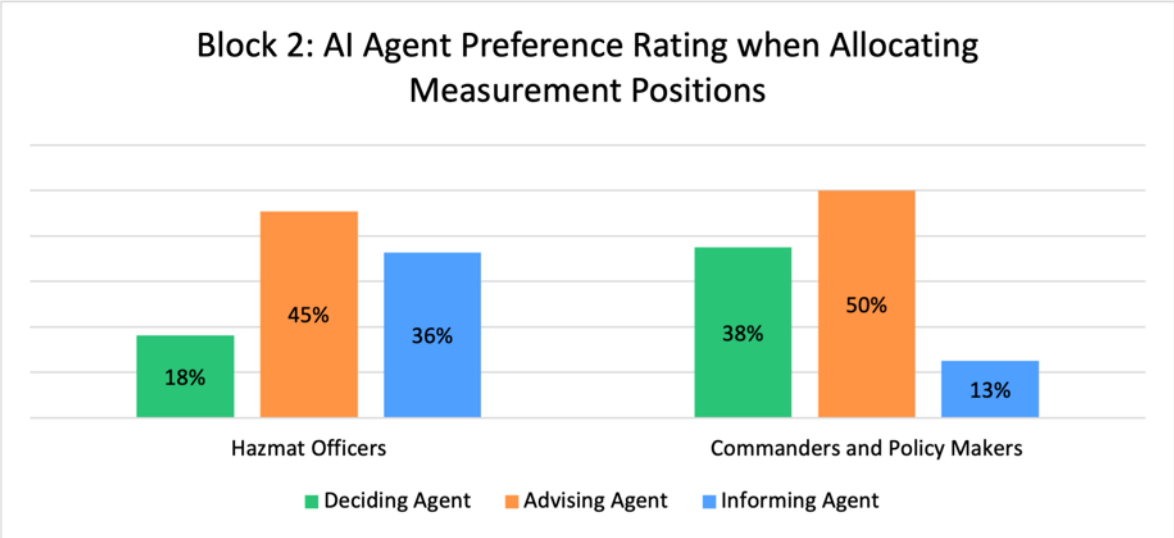


Figure 15. Block 2: AI agent preference ratings for the decision of where to allocate measurement teams, visualized for Group 1 (Hazmat Officers) and Group 2 (Commanders and Policy Officers). The percentage shown within the bars indicates how many participants within each group voted for collaborating with the according to AI agent (Deciding Agent, Advising Agent or Informing Agent). Group 1 tends to vote for the Deciding Agent less often compared to Group 2. Also, Group 1 tends to vote for the Informing Agent more often than Group 2.

6.2.7 Comments provided by the Firefighters

Comments provided by the participants that referred to the collaboration with the different AI agents are summarized here.

For an improvement of the collaboration with the AI agents, participants requested additional explanations about what data the AI agents base their calculations, conclusions and advises on.

Other participants pointed out that some of the information provided by the AI agents was more relevant for Hazmat Officers than Commanders or Policy Officers. Therefore, information should be displayed according to the position a FR holds during the mitigation of an incident.

It was also commented that the AI agents should be able to adapt and act accordingly to decisions taken by the FRs (e.g. the strategy for allocating measurement teams should be variable and adjustable to the preferences of the FR).

Further it was commented that FRs should be able to give assignments to the AI agents and they should be able to adjust them when needed.

Also, there should be the possibly to actively deny registered changes and automatically triggered recalculations (e.g. of gas cloud recalculations) as models might deviate from reality.

The *Deciding Agent* evoked mixed opinions. Two participants categorically rejected the *Deciding Agent* whereas others saw potential and remarked possible benefits. For example, one participant commented that having a functioning *Deciding Agent* would make implementation of a decision faster. Another participant suggested to link a swarm of drones to the *Deciding Agent* and let it autonomously send these drones to do measurements. The participant would dare to let the *Deciding Agent* act autonomously in that case. The swarm of drones would produce a high number of data which would increase the reliability of the model and the trust in the decision model of the *Deciding Agent*.

7 Discussion

Increasing advancements in technology pose great potential for assisting First Responders (FR) during disaster events. Still, if this technology is designed and implemented carelessly it can lead to creating more effort for FRs, inevitably putting more strain on the mitigation of an incident. Therefore, the objective of this master thesis was to perform an explorative investigation about how the collaboration between the Chemical Hazard Tool and FRs should be designed to offer suitable assistance during the mitigation of incidents that involve toxic substances.

The work was divided into two steps. First, three Team Design Patterns (*Informing Agent, Advising Agent and Deciding Agent*) were created and implemented into the Chemical Hazard Tool in form of Interaction Design Patterns (described in section 4). Second, an online survey including a low fidelity simulation was compiled to evaluate the created TDPs (see section 5). Firefighters were asked to fill out the survey and the results were presented in the previous section. In this section, the outcomes will be discussed with regard to the research questions.

7.1 Understandability and Reusability of Designed TDPs

Understandability and reusability of the created TDPs were assessed to answer the posed sub research question: "Do the created TDPs lead to a good understanding of the proposed collaboration designs and are they regarded as reusable?"

The results show that most firefighters agreed with the statement that they understood the roles and responsibilities for themselves and the AI agents. This supports the claim that TDPs facilitated communicating design choices to the stakeholders and created a common ground for discussion (van der Waa et al., 2020).

When asked whether the presented AI agents could also be used for further decisions and tasks that were not presented within the scenarios, most firefighters agreed as well. This supports the claim that TDPs provide reusable solutions (van Diggelen & Johnson, 2019).

These results were found for both assessed TDPs (TDP 1 [*Informing Agent*] and TDP 2 [*Advising Agent*]). This was expected, as all patterns were presented in the same way (see Appendix L & Appendix M).

Even though the results cannot be interpreted for the population, they do give an indication towards confirming the stated claims within the assessed study sample.

7.2 Finding a Suitable Team Design Pattern for the Human-Agent Team

In this section the implications of the results addressing the main research question will be discussed: "How should the collaboration between first responders and the Chemical Hazard Tool be designed? Which of the created Team Design Patterns is suitable to support first responders and assures adequate situation awareness??".

Assessing whether the AI agents ensured situation awareness (SA) during Block 1 of the survey showed that the SA measures were overall satisfactory. No significant difference was identified between collaborating with either of the AI agents (*Informing Agent* compared to the *Advising Agent).* Therefore, based on the SA scores no statement can be made about which collaboration design offers better support. However, the results can only be seen as an indication because due to the nature of online surveys the FRs were not exposed to the same conditions they would normally encounter during real missions.

Results from Block 1, in which the *Informing Agent* and the *Advising Agent* were assessed, indicate that FRs preferred to be supported by one of the agents rather than working without this assistance. Supporting this statement, FRs rated the assistance provided by both AI agents on average as helpful.

When trying to identify which of the three AI agents (*Informing Agent*, *Advising Agent* or *Deciding Agent*) is best suited to support the FRs, the results show that there is no "one fits all" solution. Preferences regarding which AI agent to collaborate with varied depending on the decision that had to be made and the FR that made the decision.

Results from Block 2 show that FRs were more likely to give the *Deciding Agent* permission to act autonomously for decisions that had less impact on the safety of citizens. Most FRs handed over control to the *Deciding Agent* for the decision of whether to warn team members located in a danger zone. In turn, almost none of the FRs allowed the *Deciding Agent* to act on its own when deciding whether to alert the public, in that case the preference was to collaborate with the *Advising Agent.*

Further, comparing Group 1 (Hazmat Officers) and Group 2 (On-Scene Commanders and Policy Officers) the results show that Group 1 tended to give less autonomy to the AI agent for the task of allocating measurement teams within the field. This could be related to the expertise of Hazmat Officers and their usual responsibility for this task during real life missions. In turn, Group 2 whose members are usually advised by Hazmat Officers during the mitigation of an incident involving hazardous substances, sought more assistance from the AI-agents and tended to choose the *Informing Agent* the least. Accordingly, during the simulated collaboration in Block 1, Commanders and Policy Officers (Group 2) tended to follow the recommendations of the *Advising Agent* more often than Hazmat Officers (Group 1). Again, this could be related to their respective expertise, roles and responsibilities during the mitigation of real-life incidents.

Based on these findings, neither of the presented Team Design Patterns (*Informing Agent*, *Advising Agent*, *Deciding Agent)* can be promoted as the best collaboration style nor discarded as the worst. Depending on the situation FRs are facing, their expertise and personal preference, each of the TDPs has its legitimacy. Therefore, integrating all three AI agents into the Chemical Hazard Tool and giving the FRs the option to choose seems appropriate.

Finally, feedback provided by the FRs indicates that reliability, trust, transparency and explainable AI (e.g., Neerincx et al., 2018 & De Visser et al., 2020) are important topics that have to be taken into account in the further development of the Chemical Hazard Tool.

8 Conclusion

The work performed for this master thesis shows that TDPs can be utilized to involve end-users at an early stage in the design process of a Human-Agent Team. It further shows that FRs clearly value the support of an AI agent during the mitigation of incidents, but preferences on the type of support depend on the position held within the fire brigade, the task at hand and the context of the situation. This research further contributes to the still young field of Team Design Patterns. The results found in this research will ensure that the support offered to first responders by the Chemical Hazard Tool will be tailored towards their needs and preferences.

8.1 Limitations

The findings reported in this master thesis have to be considered with a few limitations in mind that should be addressed in a follow up study. One limitation is that the evaluation was not performed with a more mature prototype and that the FRs were not exposed to the same conditions as they normally would be during a real-life mission. In these cases, FRs would receive a multitude of information through different means, which they would have to filter and prioritize. Further, on real missions, they would also have to constantly coordinate their actions with multiple other parties (e.g., paramedics, civil protection, etc.). Therefore, filling out the survey posed only a fraction of the workload FRs would normally experience during a real-life mission. A factor that influenced the decision to utilize an online survey was the difficulties of real-life meetings caused by the COVID-19 pandemic.

Another limitation is that the *Deciding Agent* was not tested within a complete scenario like the *Informing Agent* and the *Advising Agent*. Even though this was done to decrease the time FRs would need to complete the survey and therefore increase their engagement, it meant accepting the trade-off and possibly losing valuable data.

8.2 Future Research

Based on the results it is recommended to integrate all three AI agents into the Chemical Hazard Tool and to provide the option to switch from one to the other according to the preferences of the FR and the decision at hand. As this master thesis contributes to the early stages of the development and design process, further research will be needed to investigate how to ensure that the right AI agent is offering support for the right decision according to the preference of the FR. One possible solution that could be investigated is creating a higher-level Work Agreement TDP, which will govern the use of the *Informing Agent*, *Advising Agent* and *Deciding Agent*. Work Agreements are a good method to specify preferences, obligations and prohibitions (Mioch et al., 2018).

9 References

Alexander, C., Ishikawa, S., and Silverstein, M. A pattern language: towns, buildings, construction. Oxford University Press, New York, 1977.

Berggren, P., Prytz, E., Johansson, B., & Nählinder, S. (2011). The relationship between workload, teamwork, situation awareness, and performance in teams: A microworld study. *Proceedings of the Human Factors and Ergonomics Society*, 851–855. https://doi.org/10.1177/1071181311551177

Blackhurst, J. L., Gresham, J. S., & Stone, M. O. (2011). *The Autonomy Paradox*. Armed Forces Journal. http://armedforcesjournal.com/the-autonomy-paradox/

Borchers, J. O. (2001). A Pattern Approach to Interaction Design. *AI & Society*, *15*, 359–376.

Bradshaw, J. M., Feltovich, P., & Johnson, M. (2011). Human-Agent Interaction. In G. Boy (Ed.), *Handbook of Human-Machine Interaction* (pp. 293–302).

Brandweer Flevoland. (2014). Handboek Meetplanleider. Instituut Fysieke Veiligheid. https://www.ifv.nl/kennisplein/Documents/20140512-Brw-Flevoland-Handboek-Meetplanleider.pdf

Cambridge University Press. (n.d.). First Responder. In *Cambridge dictionary*. Retrieved July 15, 2021 from https://dictionary.cambridge.org/us/dictionary/english/first-responder

Chauvin, C., Lardjane, S., Morel, G., Clostermann, J. P., & Langard, B. (2013). Human and organisational factors in maritime accidents: Analysis of collisions at sea using the HFACS. *Accident Analysis and Prevention*, *59*, 26–37. https://doi.org/10.1016/j.aap.2013.05.006

Chen, J. Y. C., Procci, K., Boyce, M., Wright, J., Garcia, A., & Barnes, M. J. (2014). Situation Awareness–Based Agent Transparency. *US Army Research Laboratory*, *April*, 1–29.

Chung, E. S., Hong, J. I., James, L., Prabaker, M. K., Landay, J. A., & Liu, A. L. (2004). Development and evaluation of emerging design patterns for ubiquitous computing. *DIS2004 - Designing Interactive Systems: Across the Spectrum*, 233–242. https://doi.org/10.1145/1013115.1013148

Dekker, S. W. A., & Woods, D. D. (2002). MABA-MABA or Abracadabra? Progress on Human-Automation Co-ordination. *Cognition, Technology & Work*, *4*(4), 240–244. https://doi.org/10.1007/s101110200022

Demir, M., McNeese, N. J., & Cooke, N. J. (2017). Team situation awareness within the context of human-autonomy teaming. *Cognitive Systems Research*, *46*, 3–12. https://doi.org/10.1016/j.cogsys.2016.11.003

De Visser, E.J., Peeters, M.M.M., Jung, M.F., Kohn, S., Shaw, T.H., Pak, R. & Neerincx, M.A. (2020). Towards a Theory of Longitudinal Trust Calibration in Human–Robot Teams. *Int J of Soc Robotics* 12, 459–478. https://doi.org/10.1007/s12369-019-00596-x

Directorate-General for European Civil Protection and Humanitarian Aid Operations (ECHO) (European Commission). (2017). *Overview of natural and man-made disaster risks the European Union may face.* Luxembourg: Publications Office of the European Union https://doi.org/10.2795/861482

Doran, D., Schulz, S., & Besold, T. R. (2017). What does explainable ai really mean? A new conceptualization of perspectives. *ArXiv*.

Endsley, M. R. (1988). Design and Evaluation for Situation Awareness Enhancement. *Proceedings of the Human Factors Society, 32nd Annua*, 97–101. https://doi.org/10.1177/154193121005402801

Endsley, M. R., & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. In *Ergonomics* (Vol. 42, Issue 3). https://doi.org/10.1080/001401399185595

Gamma, E., Helm, R., Johnson, R., & Vlissides, J. Design patterns (1995): Elements of reusable object-oriented software. Addinson-Wesley Longman Inc

Gorman, J. C., Cooke, N. J., & Winner, J. L. (2006). Measuring team situation awareness in decentralized command and control environments. *Ergonomics*, *49*(12–13), 1312–1325. https://doi.org/10.1080/00140130600612788

Grigsby, S. S. (2018). *Artificial Intelligence for Advanced Human-Machine Symbiosis*. Springer International Publishing. https://doi.org/10.1007/978-3-319-91470-1

Hollnagel, E. (2007). Flight decks and free flight: Where are the system boundaries? *Applied Ergonomics*, *38*(4), 409–416. https://doi.org/10.1016/j.apergo.2007.01.010

Hollnagel, E., & Woods, D. D. (2005). *Joint Cognitive*.

Johnson, M., Bradshaw, J. M., Feltovich, P. J., Jonker, C. M., Van Riemsdijk, M. B., & Sierhuis, M. (2014). Coactive Design: Designing Support for Interdependence in Joint Activity. *Journal of Human-Robot Interaction*, *3*(1), 43. https://doi.org/10.5898/jhri.3.1.johnson

Johnson, M., & Vera, A. H. (2019). No Ai is an island: The case for teaming intelligence. *AI Magazine*, *40*(1), 16–28. https://doi.org/10.1609/aimag.v40i1.2842

Jones, D. G., & Endsley, M. R. (1996). Sources of situation awareness errors in aviation. *Aviation Space and Environmental Medicine*, *67*(6), 507–512.

Jones, W., & Endsley, M. R. (2001). A Model of Inter- and Intrateam Situational Awareness: Implications for Design, Training, and Measurement. *New Trends in Cooperative Activities: Understanding System Dynamics in Complex Environments*, *January 2001*, 46 67.

Klein, G., Feltovich, P. J., Bradshaw, J. M., & Woods, D. D. (2005). Common Ground and Coordination in Joint Activity. In *Organizational Simulation* (Issue June). https://doi.org/10.1002/0471739448.ch6

Klein, G., Woods, D. D., Bradshaw, J., Hoffman, R. R., & Feltovich, P. J. (2004). Ten Challenges for Making Automation a "Team Player" in Joint Human-Agent Activity. *IEEE Computer Society*, *December*, 91–95.

Marathe, A. R., Schaefer, K. E., Evans, A. W., & Metcalfe, J. S. (2018). Bidirectional communication for effective human-agent teaming. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10909 LNCS*, 338–350. https://doi.org/10.1007/978-3-319-91581-4_25

Matthews, M. D., & Beal, S. A. (2002). Assessing situation awareness in field training exercises. *ARI DTIC Report 1795*, *September*. http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA408560

Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. VROM. (Ministerie van Volkshuisvesting, Ruimtelijke Ordening en Milieu). (2010). *Leidraad Risico Inventarisatie Overige Ramptypen. Branchestandaarden blijvende vakbekwaamheid*. Leidraad Risico Inventarisatie - Kwetsbare objecten. Accessed: July 18, 2021. [Online]. Available: https://www.ifv.nl/kennisplein/Paginas/Leidraad-Risico-Inventarisatie-Overige-Ramptypen.aspx

Mioch, T., Peeters, M. M. M., and Neerincx, M. A. (2018): "Improving Adaptive Human-Robot Cooperation through Work Agreements". In: 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), pp. 1105–1110. (2018)

Mioch, T., Sterkenburg, R., Beuker, T. and Neerincx, M.A. (2021). *Actionable Situation Awareness: Supporting Team Decisions in Hazardous Situations*. In: Proceedings of the 18th ISCRAM Conference. Ed. By A. Adrot, R. Grace, K. Moore and C. Zobel. Blacksburg, VA, USA.

Neerincx, M. A., van Diggelen, J., & van Breda, L. (2016). Interaction design patterns for adaptive human-agent-robot teamwork in high-risk domains. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9736*, 211–220. https://doi.org/10.1007/978-3-319-40030-3_22

Neerincx M.A., van der Waa J., Kaptein F., van Diggelen J. (2018) Using Perceptual and Cognitive Explanations for Enhanced Human-Agent Team Performance. In: Harris D. (eds) Engineering Psychology and Cognitive Ergonomics. EPCE 2018. Lecture Notes in Computer Science, vol 10906. Springer, Cham. https://doi.org/10.1007/978-3-319-91122-9_18

Novak, J. D., & Cañas, A. J. (2008). The Theory Underlying Concept Maps and How to Construct Them, Technical Report IHMC CmapTools 2006-01 Rev 01-2008, Florida Institute for Human and Machine Cognition

Perez, D. R., Domingo, M. E., Perez Llopis, I. & Carvajal Rodrigo, F.J (2020). System and Architecture of an Adapted Situation Awareness Tool for First Responders. In Hughes A., McNeill F., & Zobel, C. W. (Eds.), *ISCRAM 2020 Conference Proceedings – 17th International Conference on Information Systems for Crisis Response and Management* (pp. 928–936). Blacksburg, VA (USA): Virginia Tech.

Salas, E., Cannon-Bowers, J. A., & Blickensderfer, E. L. (1993). Team performance and training research: emerging principles. *Journal of the Washington Academy of Sciences*, *83*(2), 81–106.

Schaefer, K. E., Straub, E. R., Chen, J. Y. C., Putney, J., & Evans, A. W. (2017). Communicating intent to develop shared situation awareness and engender trust in human-agent teams. *Cognitive Systems Research*, *46*, 26–39. https://doi.org/10.1016/j.cogsys.2017.02.002

Shively, J. R., Brandt, S. L., Lachter, J., Matessa, M., Sadler, G., & Battiste, H. (2016). Application of Human-Autonomy Teaming (HAT) Patterns to Reduced Crew Operations (RCO). *Springer International Publishing Switzerland*, *9736*, VII. https://doi.org/10.1007/978-3-319-40030-3

Shu, Y., & Furuta, K. (2005). An inference method of team situation awareness based on mutual awareness. *Cognition, Technology and Work, 7*(4), 272–287. https://doi.org/10.1007/s10111-005-0012-x

Van der Aalst, W. M.,. Ter Hofstede, A. H, Kiepuszewski, B. & A. P. Barros (2003). *Workflow patterns*, Distrib. Parallel Databases, vol. 14, no. 1, pp. 5–51, 2003.

Van der Waa, J., van Diggelen, J., Cavalcante Siebert, L., Neerincx, M., & Jonker, C. (2020). Allocation of moral decision-making in human-agent teams: a pattern approach. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12187 LNAI*, 203–220. https://doi.org/10.1007/978-3-030-49183-3_16

Van der Waa J., Verdult S., van den Bosch K., van Diggelen J., Haije T., van der Stigchel B., & Cocu I. (2021). *Moral Decision Making in Human-Agent Teams: Human Control and the Role of Explanations.* Frontiers in Robotics and AI, 8, 143. https://doi.org/10.3389/frobt.2021.640647

Van Diggelen, J., & Johnson, M. (2019). Team design patterns. *HAI 2019 - Proceedings of the 7th International Conference on Human-Agent Interaction*, 118–126. https://doi.org/10.1145/3349537.3351892

Van Diggelen, J., Neerincx, M., Peeters, M., & Schraagen, J. M. (2019). Developing Effective and Resilient Human-Agent Teamwork Using Team Design Patterns. *IEEE Intelligent Systems*, *34*(2), 15–24. https://doi.org/10.1109/MIS.2018.2886671

Van Stijn, J. (2020). *Moral Decision-Making in Medical Hybrid Intelligent Systems A Team Design Patterns Approach to the Bias Mitigation and Data*. Vrije Universiteit Amsterdam. Master's Thesis.

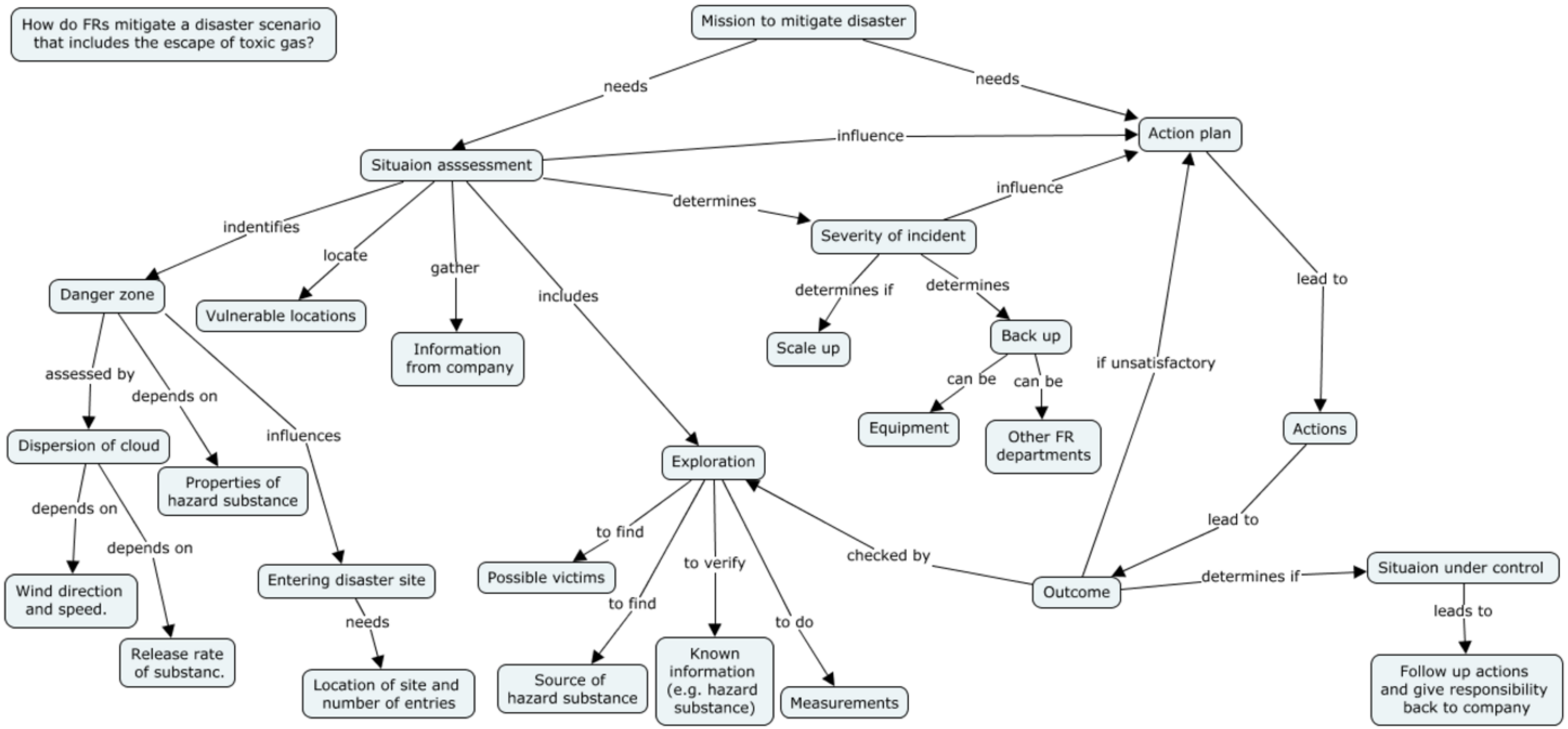
Van Stijn, J., Neerincx, M. A., Annette, T., & Vethman, S. (2021). *Team Design Patterns for Moral Decisions in Hybrid Intelligent Systems: A Case Study of Bias Mitigation*.

Van Welie, M., van der Veer, G. C., & Eliëns, A. (2001). Patterns as Tools for User Interface Design. *Tools for Working with Guidelines*, 313–324. https://doi.org/10.1007/978-1-4471-0279-3_3

10 Appendix

10.1 Appendix A

Concept Map mission evolvement



10.2 Appendix B

Goal-Decision- Required Information Table

Goal	Decision	Information needed
Phase 1		
Enter side safely.	- How to enter incident site?	- Location of incident - Number and location of possible entries. - Wind direction and speed - Release rate of substance - Temperature outside
Phase 2		
Explore side safely (get an overview of the situation).	- Can the team members be sent out or do robots/drones have to be send? - What equipment is needed? (Hazard suit?) - Where to explore?	- Properties of hazard substance - Expected affected area - Equipment available
Evaluate severity of incident.	- How big is the range of the affected area? - Is back up needed?	- Number of complaints received by the public-safety answering point. - Vulnerable spots (e.g. hospital, residential areas, etc.) - Size and complexity of incident.
Plan and carry out actions to mitigate incident.	- How should possible victims be saved? - Do possible victims need first aid? - How to close a possible leak? - Can the company block pipes? - Can the cloud be diffused further? - Do vulnerable spots need to be warned?	- Number of victims - Pressure of leak. - Properties of hazard substance. - Resources company has. - Estimation of how long the process of stopping the source will take
Verify outcome of taken actions.	- Is the source of danger neutralized?	- Information from measurements
Phase 3		

Wrap up mission and hand back responsibility to company.	- Is there a need to disinfect people and equipment? - What instructions does the company need to proceed? - What other measures have to be taken before the responsibility can be given back to the company?	- Knowledge about who was in the danger zone. - Knowledge about how to keep source neutralized.

10.3 Appendix C

Use Case: Chlorine Scenario

TITLE	**CHLORINE INCIDENT**
OBJECTIVE	Whole mission – mitigation of an incident
ACTORS	
PRE-CONDITION	Alarm is received
POST-CONDITION	The danger is mitigated and the situation is under control
ACTION SEQUENCE	1. It is May, a Tuesday, 1 am. (non-corona times). An alarm comes in. A train crashed on a railroad that is used to transport dangerous goods. The train was transporting chlorine. A yellow/greenish cloud is escaping and spreading. 2. Following information is provided: • Wind speed: 5 m/s • Temperature is 11°C • Wind direction: Southwest 3. The firefighting crew is getting in the car. 4. The commander opens the Chemical Hazard Tool (CBRN tool) on the tablet to get a first impression of the situation. 5. After consulting the gas cloud distribution, a safe route to get to the location is chosen. 6. The firefighting team arrives at the scene 7. They receive the information that several youths were sitting in the field when they first saw the train pass and then heard the accident. Shortly after they started to smell an intense scent, were getting irritated eyes and had trouble breathing. They managed to run away from the gas but were still feeling the stinginess of the gas they inhaled. 8. The ambulance is taking care of them. 9. No trace of the train driver yet. 10. It appears to be a continuous leak as you can see that the gas is growing and expanding steadily. 11. ProRail forwarded the information that there were no other gases than chlorin transported in that train.

12. The firefighter looks at the tablet and chooses to see vulnerable locations.
13. Using the Chemical Hazard Tool, a decision is made towards taking measures in the affected area.
14. One firefighting team is sent out to put up a water screen that disperses the chlorine.
15. A decontamination team is requested.
16. The incident is scaled up to GRIP 3.
17. Three measurement teams are sent out to the city downwind to do measurements.
18. The firefighter assigns measurement locations to the different teams in collaboration with the Chemical Hazard Tool.
19. The train driver has been located unconscious close by the train.
20. A team in protective gear is in the process of recovering the train driver.
21. The disinfection team has arrived.
22. A change in wind direction is recognized by the Chemical Hazard Tool. A measurement team is located within a danger area. A warning is sent out.
23. The leak has been closed.
24. Now measurements in the affected area will be made until no more toxic substances are detected.

10.4 Appendix D

Understandability, Reusability and Helpfulness of the *Informing Agent* and the *Advising Agent*

How much do you agree with the following statements about the presented intelligent system and the described collaboration?

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I understood the roles and responsibilities for the system and myself.	○	○	○	○	○
An Information Providing Intelligent System could also be used for additional decisions that were not shown during the mission.	○	○	○	○	○
The support of the intelligent system was helpful.	○	○	○	○	○

What would you change in the way the intelligent system provided support?

10.5 Appendix E

Scenario Related Questions

In this section you are asked about how you perceived the scenario.

I have experienced similar situations in the past.	Never	Sometimes	Often	Usually	Always
The presented scenarios and decisions were realistic.	Strongly disagree	Somewhat disagree	Neither	Somewhat agree	Strongly agree

10.6 Appendix F

Subjective Situation Awareness Rating Scales adapted from the Mission Awareness Rating Scale (MARS, Matthews & Beal, 2002).

I was able to get a good understanding of what was going on during the scenario.	Strongly disagree	Somewhat disagree	Neither	Somewhat agree	Strongly agree
It was easy to identify mission-critical information during the scenario.	Strongly disagree	Somewhat disagree	Neither	Somewhat agree	Strongly agree
It was easy to make mission relevant decisions.	Strongly disagree	Somewhat disagree	Neither	Somewhat agree	Strongly agree
It was easy to predict what was about to occur.	Strongly disagree	Somewhat disagree	Neither	Somewhat agree	Strongly agree

10.7 Appendix G

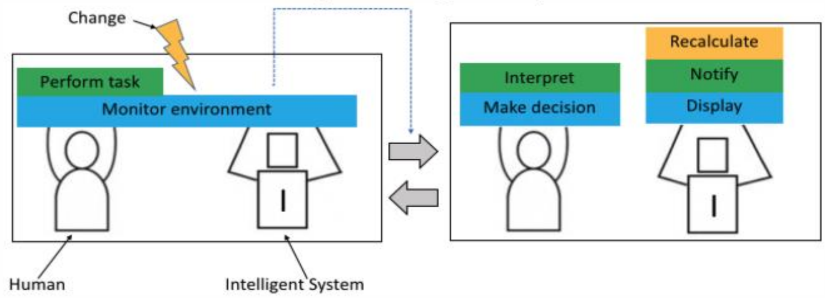
Preference Ratings at the End of the Block 1

assistance UNIVERSITEIT TWENTE. TNO innovation for life

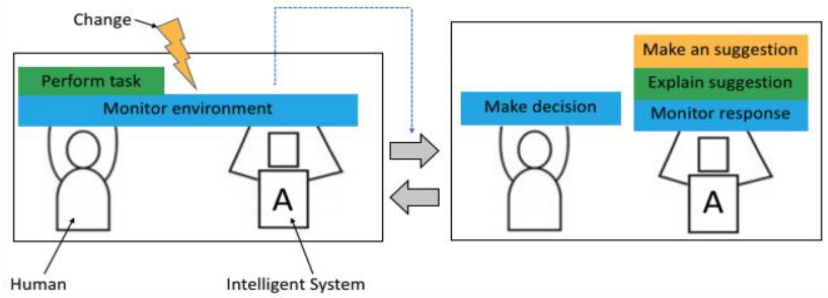
English

This section is about your preference towards the Information Providing Intelligent System versus the Advising Intelligent System.

Information Providing Intelligent System:



Advising Intelligent System:



Below is a list of decisions you took during the mission. Please indicate which system you preferred assisting you during each decision.

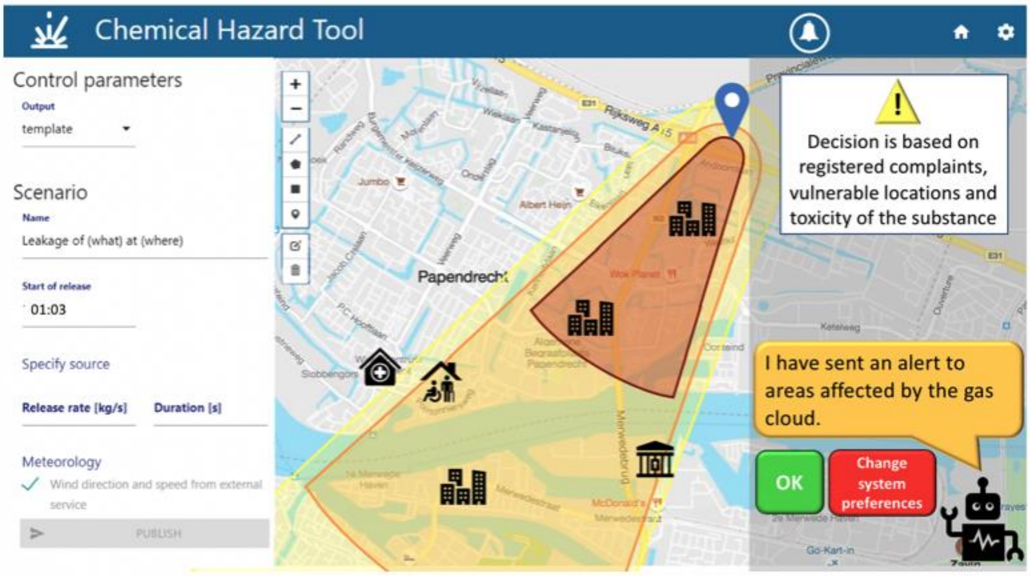
	Information Providing Intelligent System	Advising Intelligent System	Both are alright	Neither
Deciding on a route to get to the location.	○	○	○	○
Deciding whether to alert the public.	○	○	○	○
Deciding where to send teams to do measurements.	○	○	○	○
Deciding whether to warn a team about being in a danger zone.	○	○	○	○

10.8 Appendix H

Acceptance and Preference Rating during Block 2

English

The Deciding Intelligent System initiates the process to alert the public about the escaping gas. The information will include recommendations to stay inside, close windows, etc.



Would you give the intelligent system permission to (generally) decide on its own about alerting the public?

- ◯ Yes
- ◯ No, change to Information Providing Intelligent system.
- ◯ No, change to Advising Intelligent System.

← →

10.9 Appendix I

Email text for Participant Acquisition.

(English below)

Beste brandweerman,

Voor het ASSISTANCE-project werken we bij TNO aan een systeem dat u ondersteunt bij het beperken van calamiteiten waarbij giftige stoffen betrokken zijn. Het systeem kan gegevens van verschillende sensoren ontvangen en berekeningen maken.

Met dit onderzoek willen we kijken hoe de samenwerking tussen uw en het systeem kan worden vormgegeven. In de evaluatie komt u verschillende incidenten tegen die u samen met verschillende intelligente systemen oplost.

De evaluatie duurt ongeveer 30 minuten.

Het is het beste om dit op een laptop of computer te doen, omdat u schermafbeeldingen te zien krijgt die misschien moeilijker te zien zijn op een smartphone.

De evaluatie is beschikbaar in het Engels en Nederlands. Rechtsboven kunt u kiezen welke taal uw voorkeur heeft. (Excuses bij voorbaat als u termen tegenkomt die verwijzen naar termen of procedures voor brandbestrijding die niet helemaal juist zijn).

Uw deelname draagt bij aan het ontwerpen van een systeem dat bij u past. Daarom nodigen we u uit om deel te nemen door op onderstaande link te klikken.

Link

Mocht u nog vragen hebben, neem dan contact op met tatjana.beuker@tno.nl.

----English version-------

Dear firefighter,

For the ASSISTANCE project we at TNO are working on designing a system that supports you during the mitigation of emergency incidents which involve toxic substances. The system can receive data from various sensors and make calculations.

With this research we want to look at how the collaboration between you and the system could be designed. In the evaluation you will encounter different incidents which you will solve in collaboration with different intelligent systems.

The survey takes approximately 30 minutes.

It is best to do it on a laptop or computer because you will be shown screenshots that might be more difficult to see on a smartphone.

The survey is available in English and Dutch. On the top right you can choose which language you prefer. (Apologies upfront if you encounter terms referring to firefighting vocabulary or procedures that is not completely accurate).

Your participation will contribute to designing a system that fits your needs. Therefore, we invite you to participate by clicking on the link below.

Link

If you have any further questions, please contact tatjana.beuker@tno.nl.

10.10 Appendix J

Informed Consent

English

Informed Consent

Evaluating Teaming Patterns Between Firefighters and Intelligent Systems

In this evaluation you will cooperate with different types of intelligent systems to solve incidents which involve toxic substances.

The information and data you provide cannot be traced back to you. Therefore, your privacy is preserved. The data will be erased 5 years after completing the research.

The evaluation will take approximately **30 minutes.**
You can participate if you are **18 years or older** and an active **member of the fire brigade**.

If you want more information click here.

By checking the box below, you consent to the following statements:

- I have read and understood the information regarding the research.
- The intentions of the research and the followed approach have been explained to my satisfaction.
- I have had enough time to think about participating.
- I know that my participation in this study is completely voluntary and that I can withdraw my consent at any time without giving any reason.
- I give permission to process the data I provide (including age, position within the fire brigade and years of experience) for the purposes of this research.
- I give permission to reuse my research data for future research in the described research area with the condition that these are coded in such a way that they cannot be traced back to me as a person.
- I give permission to the storing of the data and that authorized members of the research team and authorized inspectors can access the data.

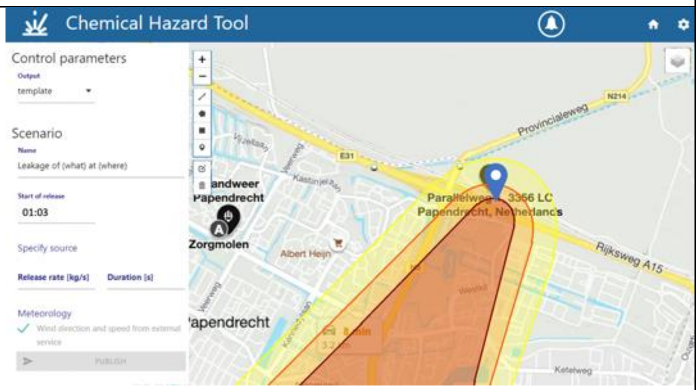
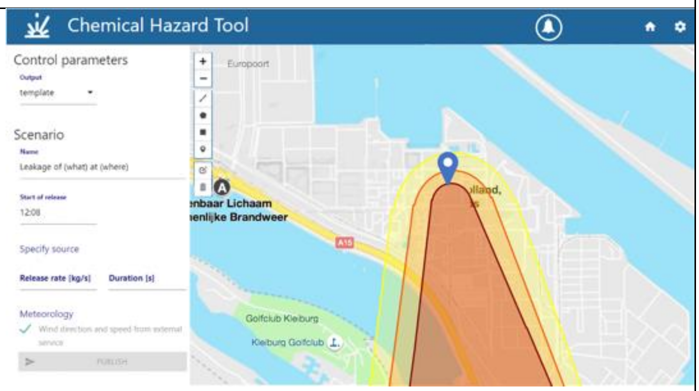
☐ I consent with the statements above.

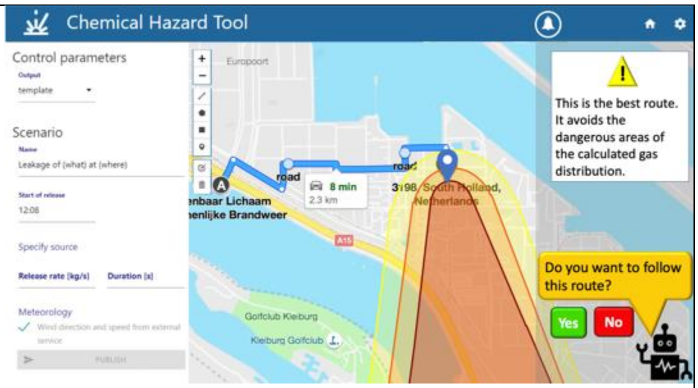
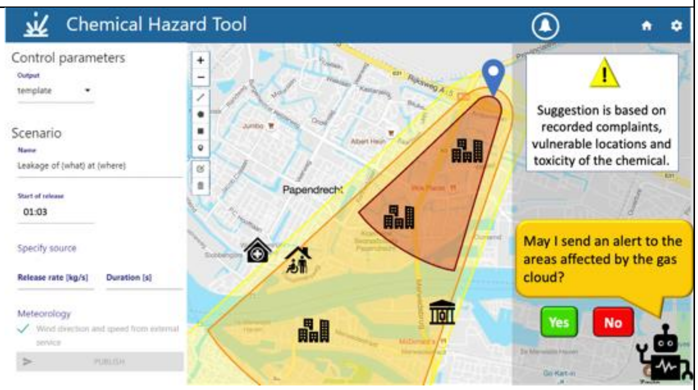
If you have any further questions or concerns, please contact:
t.beuker@student.utwente.nl

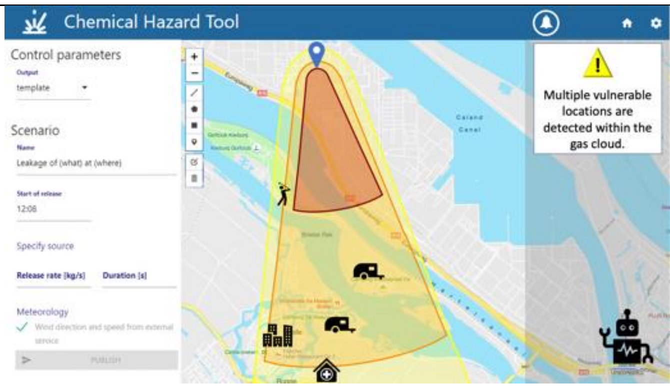
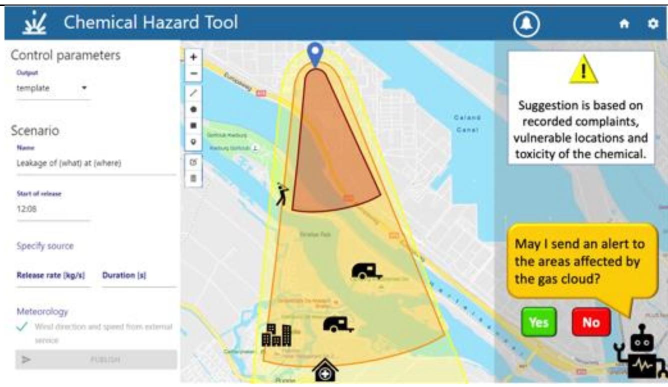
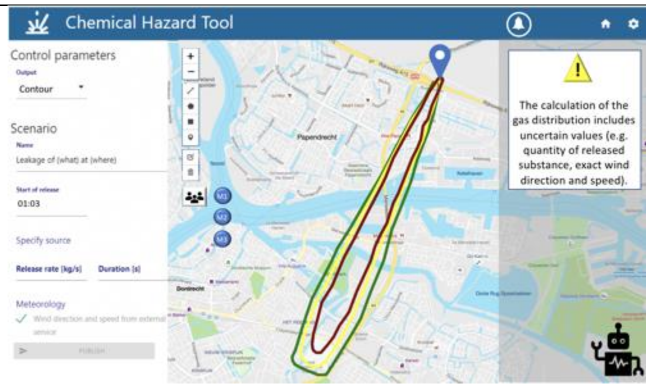
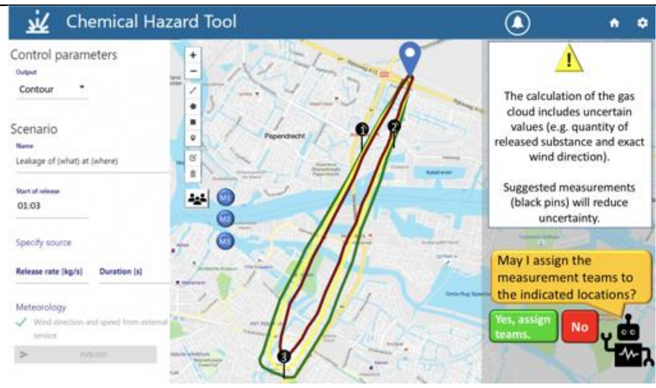
This research project has been reviewed and approved by the TNO Ethics Committee and the BMS Ethics Committee.

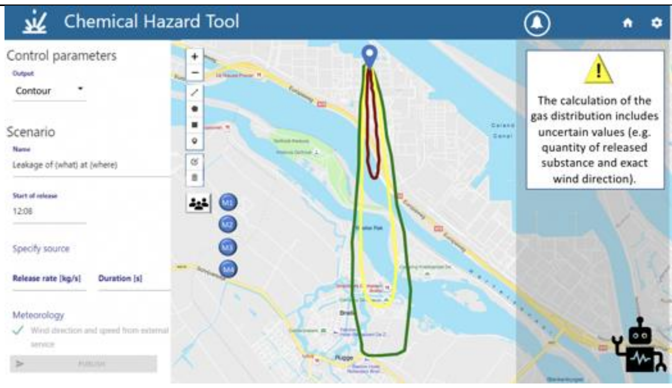
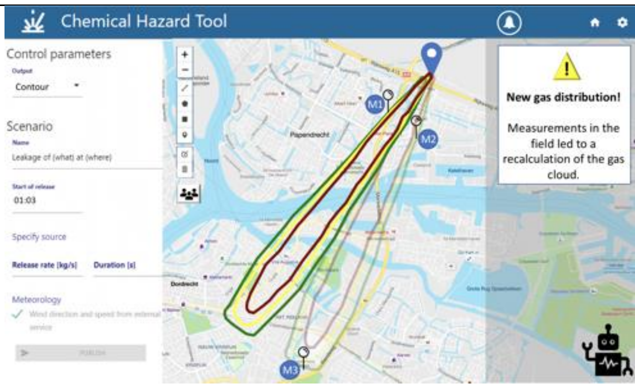
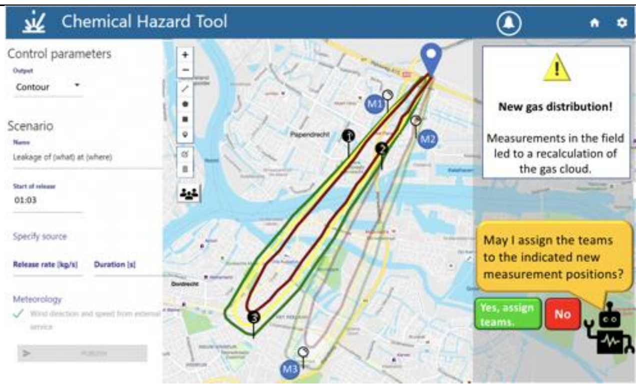
10.11 Appendix K

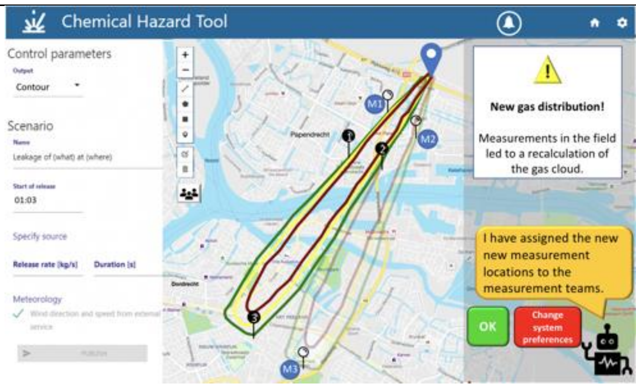
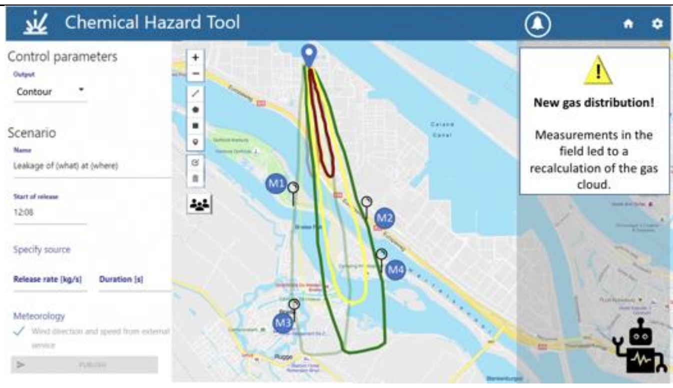
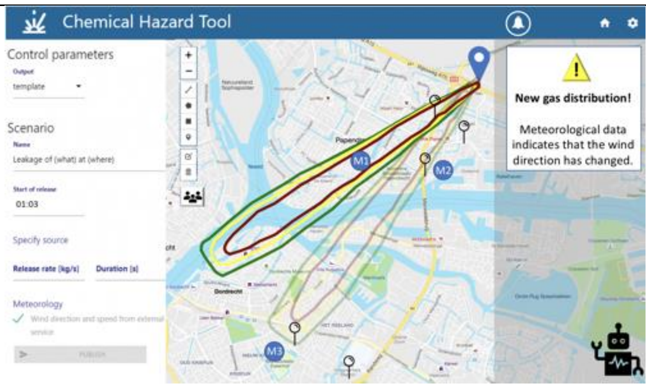
Screenshots of all visualized interactions with the three AI agents according to scenario and decision point.

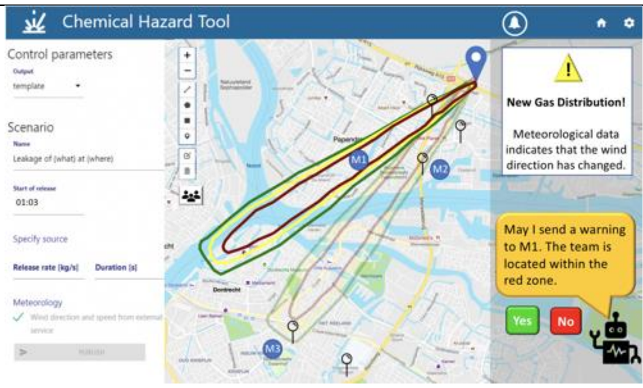
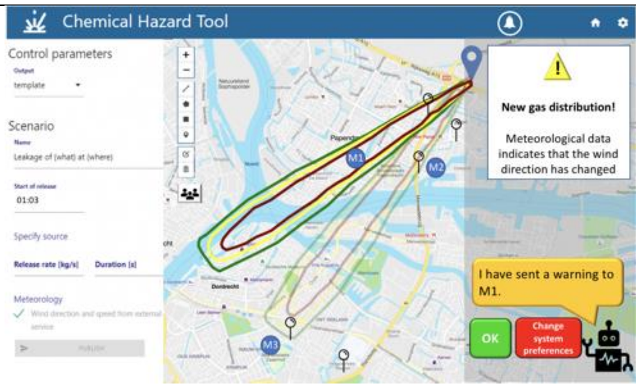
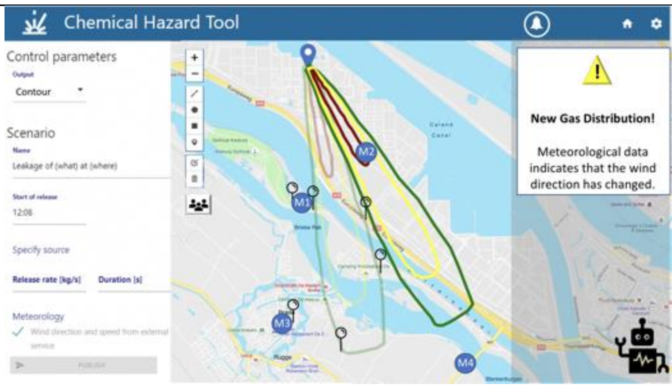
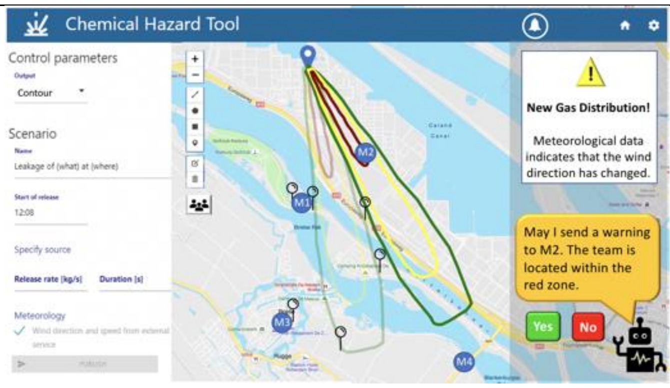
Decision/ Task	Scenario	AI Agent	Visualization
Deciding on a route to the incident location.	Chlorine Scenario	Informing Agent	
		Advising Agent	
	Sulfur Dioxide Scenario	Informing Agent	

		Advising Agent	
Deciding whether to alert the public	Chlorine Scenario	Informing Agent	
		Advising Agent	
		Deciding Agent	

	Sulfur Dioxide Scenario	Informing Agent	
		Advising Agent	
Deciding where to do measurements	Chlorine Scenario	Informing Agent	
		Advising Agent	

	Sulfur Dioxide Scenario	Informing Agent	
		Advising Agent	
Measurement nr.2.	Chlorine Scenario	Informing Agent	
		Advising Agent	

<table>
<tr><td></td><td></td><td>Deciding Agent</td><td>
</td></tr>
<tr><td></td><td>Sulfur Dioxide Scenario</td><td>Informing Agent</td><td>
</td></tr>
<tr><td></td><td></td><td>Advising Agent</td><td>
</td></tr>
<tr><td>Warning Team Member</td><td>Chlorine Scenario</td><td>Informing Agent</td><td>
</td></tr>
</table>

		Advising Agent	
		Deciding Agent	
	Sulfur Dioxide Scenario	Informing Agent	
		Advising Agent	

10.12 Appendix L

Introducing the *Informing Agent* to the Participants.

English

Information Providing Intelligent System

You will work together with an intelligent system that receives information about:

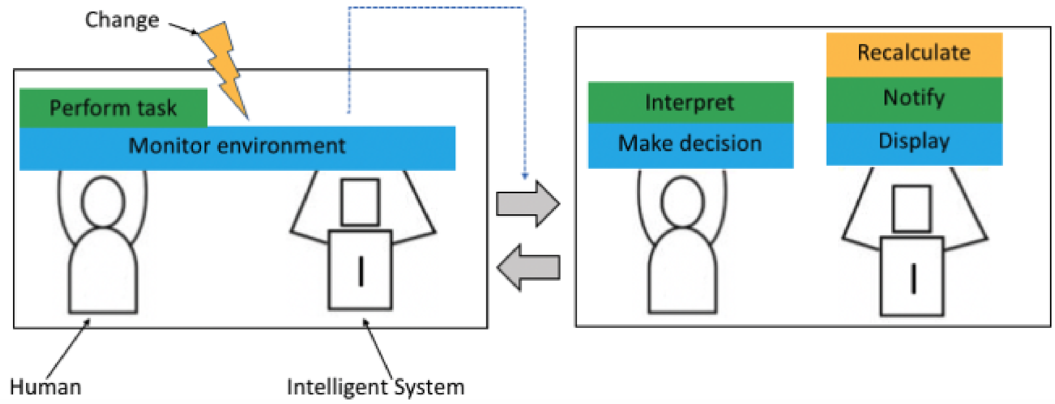
- Weather conditions,
- GPS location,
- Vulnerable locations,
- Gas measurements.

It can use this data to calculate and visualize an expected gas cloud.

The system has the responsibility to **inform** you.
You interpret the information and make a decision about what actions to take.

The information the system provides includes a visualization of received data and possible notifications about changes.

The image below shows how you will work together. The round figure represents you and the rectangular figure represents the intelligent system. Both figures hold up blocks with tasks that each performs. The flash indicates a change (e.g. wind direction change) which initiates the transition to the second frame. The intelligent system notices the change, recalculates the situation and notifies you. Based on the received information you make a decision about how to deal with the new situation. Once a decision is made you transition back to the first frame.



10.13 Appendix M

Introducing the *Advising Agent* to the Participants.

assistance UNIVERSITEIT TWENTE. TNO innovation for life

English

Advising Intelligent System

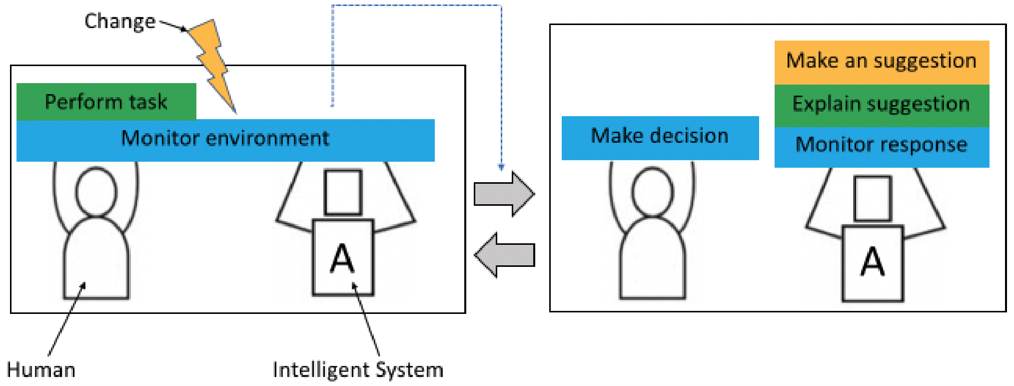
You will work together with an intelligent system that receives information about:

- Weather conditions,
- GPS location,
- Vulnerable locations,
- Gas measurements.

It can use this data to calculate and visualize an expected gas cloud.

The system has the responsibility to inform you about changes in the environment and **advise** you on how to adapt to them.
You consider the advice and decide whether to accept it or not.

The image below shows how you will work together. The round figure represents you and the rectangular figure represents the intelligent system. Both figures hold up blocks with tasks that each performs. The flash indicates a change (e.g. wind direction change) which initiates the transition to the second frame. The intelligent system notices the change, reassess the situation and makes a suggestion to assist you in your decision-making process. You then can accept or decline it. Once a decision has been made you transition back to the first frame.



10.14 Appendix N

Introducing the *Deciding Agent* to the Participants.

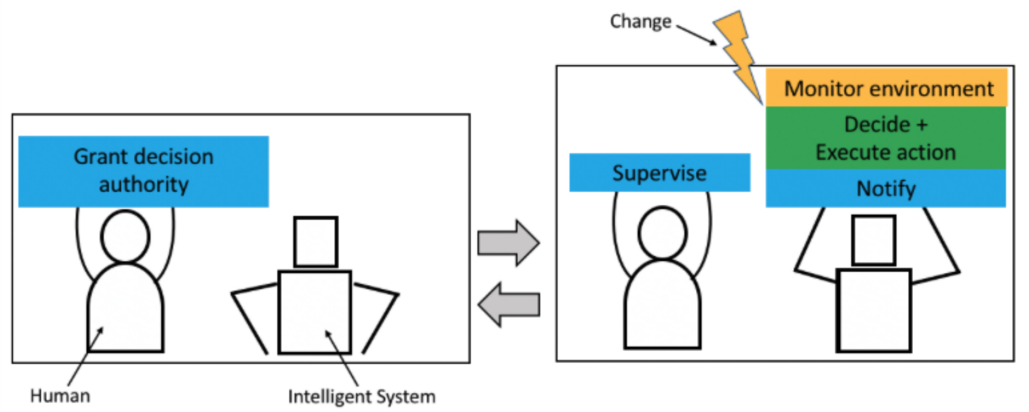
assistance UNIVERSITEIT TWENTE. TNO innovation for life

English

You have been introduced to two intelligent systems. One that only provides you with information and the other one that additionally gives you advice.
We will now introduce you to a third intelligent system, that is allowed to decide on its own. It can then act upon the decisions it has made.

For example, instead of getting first your permission on whether to warn team members about being located within a dangerous area it will straight away send the message to that team and afterwards inform you about the action it has performed.

On the next page you will see some of the decisions that were taken during the chlorine incident, but this time you will see how the interaction would have been with a **Deciding Intelligent System.**



→