Design for Integrating Explainable AI for Dynamic Risk Prediction in Ambulance IT-systems

Master's thesis in Interaction Design

David Wallsten & Gregory Axton

Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2023

Master's thesis 2023

Design for Integrating Explainable AI for Dynamic Risk Prediction in Ambulance IT-systems

David Wallsten

Gregory Axton

UNIVERSITY OF GOTHENBURG

CHALMERS UNIVERSITY OF TECHNOLOGY

Department of Computer Science and Engineering
Chalmers University of Technology
Gothenburg, Sweden 2023

Design for Integrating Explainable AI for Dynamic Risk Prediction in Ambulance IT-systems

David Wallsten & Gregory Axton

Supervisor: Mikael Wiberg, Department: Interaction Design and Software Engineering, Computer Science and Engineering
co-Supervisor: Thommy Eriksson, Department: Interaction Design and Software Engineering, Computer Science and Engineering
Supervisor: Stefan Candefjord, Care@Distance
co-Supervisor: Eunji Lee, Care@Distance
Examiner: Palle Dahlstedt, Department: Interaction Design and Software Engineering, Computer Science and Engineering

Master's Thesis 2023
Department of Computer Science and Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Printed by Chalmers Reproservice
Gothenburg, Sweden 2023

Design for Integrating Explainable AI for Dynamic Risk Prediction in Ambulance IT-systems

David Wallsten & Gregory Axton
Department of Computer Science and Engineering
Chalmers University of Technology

Abstract

Demographic changes in the West with an increasingly elderly population puts stress on current healthcare systems. New technologies are necessary to secure patient safety. AI development shows great promise in improving care, but communicating AI decisions requires more research. In this study, a prototype of eXplainable AI (XAI) was designed for an ambulance IT system, based on an AI model for risk prediction of severe trauma to be used by Emergency Medical Services (EMS) clinicians.

Knowledge was gathered for the design through ethnography, expert interviews, and a literature review. Then several ideas developed through stages of prototyping and were verified by experts in prehospital healthcare. Finally, a high-fidelity prototype was evaluated by the EMS clinicians.

The design was then evaluated by seven EMS clinicians. They thought that XAI was necessary for them to trust the prediction. They make the final decision, and if they can't base it on specific parameters, they feel they can't make a proper judgement. In addition, the design helped in reminding EMS clinicians of things they might have missed. If given a prediction from the AI that was different from their own, it might cause them to think more about their decision, moving it away from the normally relatively automatic process and likely reducing the risk of bias.

While focused on trauma, the design should be applicable to other AI models. Current models for risk prediction in ambulances have so far not seen a big benefit from using artificial neural networks (ANN) compared to more transparent models. This study can help guide the future development of AI for prehospital healthcare and give insights into the potential benefits and implications of its implementation. The report also explores the ethical implications, the complexity of the ambulance work environment and possible implications for cognitive decision processes.

Keywords: Computer, science, computer science, engineering, project, thesis.

Acknowledgements

We would like to thank everyone at the Care@Distance research group for staying interested. Especially Stefan Candefjord and Eunji Lee, who have worked as our mentors in the group, guiding us extensively throughout the process and helping us with contacts and inspiration. Also, a special thanks to Anna Bakidou, the primary developer of the AI model this research is partially based on, Beng-Arne Sjöqvist for his support in understanding the current market and the future and Mattias Seth for his knowledge contribution to interoperability.

We want to thank PICTA, especially Elin Maxstad, who did a similar project last year. With her report, this project reached nearly as far as it did the contacts she provided. We are also grateful for Andreas Dehre's contribution of knowledge and contacts in Gothenburg's ambulance services.

Also, we would like to express our gratitude to Paratus and Jonas Borgström, who welcomed us to visit the company and gain insights about their current systems. We are very thankful to Stefan Jönsson at VGR for his expertise and for informing us of the potential future of healthcare at VGR.

Lastly, thank you to all the ambulance personnel who have willingly given us their time and effort. The work would have been almost worthless without your contribution.

Everyone we met has shown excitement for our project, giving us the energy to push forward. There is much kindness and engagement in this field, with people working hard to save lives in many ways.

David Wallsten & Gregory Axton, Gothenburg, 2023-06-18

Contents

List of Figures

List of Tables

1

Introduction

1.1 Background

Healthcare in Europe is facing a major crisis with an ageing population. From 2007 to 2022, the median age has risen from 41.9 years 2007 to 44.1 years 2022(Eurostat, 2022) Figure 1.1. The same applies to Sweden, with the age group over 90 increasing the most by percentage (SCB, 2022). This increases the demand for healthcare systems with limited resources, creating a need for new solutions. The Swedish Government has a goal to be leading in e-Health by 2025 because of this reason (Blix & Levay, 2018). It creates demands on simple e-health solutions like SMS reminders but also more robust data management (e.g. big data solutions) as well as advanced decision support.

While necessary, the digital transformation of the Swedish healthcare system faces many challenges. It's important to take into account the usability and the work environment of healthcare professionals when digitising healthcare. Otherwise, it risks increasing stress levels of healthcare professionals and decreasing patient safety, as stated by Sveriges Läkarförbund (The Swedish Medical Association, 2021). This is just as relevant for IT systems in prehospital care and care outside the hospital.

A research group at Chalmers University, Care@Distance – Remote and Prehospital Digital Health, are investigating multiple new technologies with the potential to improve prehospital care. They include AI-driven decision support (Candefjod et al., 2021), video-supported assessments in ambulances (Vicente et al., 2021), and voice-controlled augmented reality interfaces (Kemppainen, 2022). The hope is to make decisions faster, more efficiently and safer, often using big data tools with wide combinations of data. It can be both structured, such as tabular data, and unstructured such as long-form text or images. However, further user experience(UX) research is needed to fully integrate these technologies into the workflow of ambulance workers and determine the most effective means of presenting the results of the decision support systems. The field of explainable AI (XAI), in particular, is very new, with big needs for user studies (Antoniadi, et.al., 2021). XAI lacks a proper technical definition, but key elements are interpretability, how much the model can be understood, transparency, about how the model works, and explainability, giving insight into the reasons for the AIs' decisions (Antoniadi, et.al., 2021).

Currently, ambulance care uses different triage decision support systems, such as

Rapid Emergency Triage and Treatment System (RETTS) or West Coast System for Triage (WEST), a kind of checklist based on vital parameters, patient information and warning symptoms. The problem with these systems is that they can be imprecise. In the case of children, RETTS can lead to an over-triage of around 30% and the same level of under-triage (Magnusson, 2021). Over-triage is when a case is rated more severe than it is, leading to resource drain. Under-triage is when a case is rated as less severe, leading to a high risk to patient safety. AI decision support systems have the possibility to have higher precision than current systems. A particular AI model is under development by Anna Bakidou et. al. (2023) at Care@Distance. It is a general model for trauma triage, assessment and prediction for urgency and severity. Bakidou's model uses data usually collected at a hospital to make predictions for the ambulance. Thus, a novel design is needed.

With Bakidou's model as a base for the design, this project aims to investigate how data from AI models could be presented in a way that fits into the ambulance workflow. More importantly, the question of how to display the AI model's prediction is important to build trust and understanding with the ambulance workers, who are ultimately responsible for making the final decision. And could this information be displayed earlier than current systems, leading to quicker decision-making?

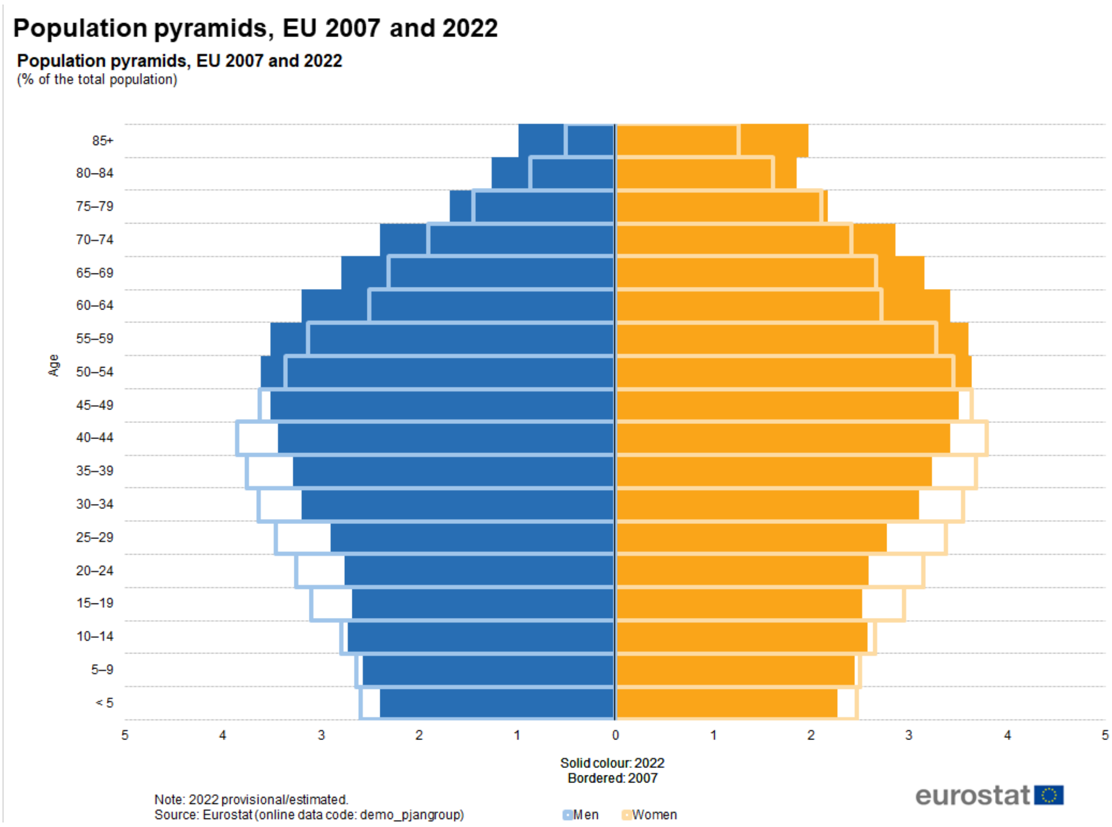


Figure 1.1: Population pyramids, EU 2007 and 2022 (Eurostat, 2022)

1.2 Stakeholders

Care@Distanc ASAPs
Care@Distance ASAP (Acute Support Assessment and Prioritising) is a research group based at Chalmers University with a vision to increase decision precision, limiting errors in assessment, prioritisation and handling while also reducing time to make decisions. Aims to improve remote and prehospital care using data fusion, clinical decision support, AI/ML (Machine learning), telemedicine, and innovative user interaction. Care@Distance ASAP is the main stakeholder for this project. It provided valuable resources and insights within the healthcare sector input to drive this project forward. The expectation is to gain influence by this study as a part of the more prominent research. It may raise a discussion on the design aspect of future systems and ongoing research.

PICTA
PICTA is a prehospital innovation arena funded by the region Västra Götaland (VGR) as a part of Lindholmen Science Park. Their goal is to be a platform for research sharing in the prehospital setting, to make individual projects inform each other, driving progress forward. It is a collaborative platform for health care, academia, and business organisations. Different governmental foundations, such as Vinnova, fund projects run or supervised by PICTA. PICTA is not directly involved in the development of this project. Elin Maxstad introduced us to her previous report and provided us with the necessary information and contacts to drive this project forward. Andreas Dehre, a former nurse and current teacher of ambulance personnel, has provided us with his expertise and contacts as well, including inviting us to the ambulance station in Gothenburg. PICTA has solid knowledge and experience within prehospital settings and the implementation of new technology in this field. PICTA does not expect anything in terms of deliverables and is involved solely by self-interest, and will act as continuous support throughout the project.

Västra Götaland Region (VGR)
The Västra Götaland Region (VGR) is the governmental organisation that funds healthcare in Västra Götaland including the city of Gothenburg. VGR is planning to implement a new comprehensive healthcare system for all healthcare in VGR, including prehospital care. The latter haven't started development yet. To narrow the scope of this project, it is focused on VGR instead of other regions, and basing its design on the capabilities of the new IT system. Thus, this project could work as a source of influence for any future ambulance system when it starts development.

Hospital
The emergency department (ED) at hospitals in Sweden are the departments that have direct contact with the ambulance. ED collects valuable information such as ambulance action and data like transportation, first medical treatment, patient transfer and mission confirmation. In VGR, most of the data is given to the ED through the delivery of a paper journal. In Halland, most of the data updated in the ambulance are displayed in real-time at the ED station. The ambulance system needs to cooperate with the system at the hospital, including phone calls directly through the ambulance system, and synchronise the data that need to be reported to

a doctor; therefore, the hospital needs to understand AI risk prediction and general UI to gain a better workflow between them.

Third-party service developers
In the prehospital setting in Sweden, there are multiple IT system providers. The biggest is CSAM with Paratus and Ortivus with Mobimed, according to our interview with Bengt-Arne Sjöqvist, a healthcare researcher and industry veteran. The project aims for the design to be able to be integrated into current systems, thus it is important to take these companies into consideration. The design may then work as an inspiration for future implementation of AI decision support.

1.3 Users

Primary users
The primary user is the EMS personnel who intervenes in EMS services. In this study, We have divided the user into two groups: the Primary user and the secondary user, based on who may have direct contact with the system. The primary users include EMS nurses (usually taking decisions), EMS paramedics, EMTs, EMS volunteers, EMS students, and EMS trainees. The primary user generally works in pairs that vary in age, gender, experience and skill sets.

Secondary users
The secondary user consists of those who may display information indirectly through the system under various circumstances, e.i. The EMS personnel share some info with secondary users according to their needs. This may include the patient, the patient's family or relative, and the caretaker. In addition, the secondary users may be described as witnesses and have less direct contact with the system.

1.4 Aim

This study aims to investigate a possible design for a future IT-This study aims to investigate a possible design for a future IT-system with an integrated dynamic risk prediction AI decision support tool. It will be based on an AI model trauma risk prediction but should be applicable to other conditions and AI models. It aims to act as inspiration for those working with developing the models, to get an idea about how their models could be applied and how medical professionals might react to them. It also aims to act as an inspiration for the healthcare services and industry, to help them in generating ideas on how to integrate AI into their existing systems and get inspiration for the interaction design of their general ambulance IT system. The design should take cognitive resources into account, be based on usability research and be adapted to the ever-changing work environment of ambulance personnel.

1.5 Deliverable

The project is located in the city of Gothenburg, a part of VGR. Their current work towards a new digital interface and the ease of access to local expertise and testers makes this project focused towards VGR. At the same time, the project takes current systems used nationally and internationally into account, likely making findings applicable to a much wider market.

The deliverables are:

An interactive interface for explainable AI(XAI) output A wider interactive prototype interface, imagining future utilisation of big data to increase speed and ease of use. Needs to be able to be adaptable to any medical emergency where an ambulance might be involved

1.6 Limitations

The project will be focused on the VGR region but should be adaptable to other areas. It will take into account currently available technology but might adapt them in ways currently not done in the prehospital field. This places the project with a possible implementation of an estimated 5-8 years in the future.

Primary users, meaning ambulance personnel, will be the focus of the report. The system however will have a heavy impact on secondary users, meaning patients, relatives and bystanders. Having that in mind is important. In addition, ambulance personnel might need to show the display to the secondary users in case they want to explain something, or, it might be placed in a location where secondary users might look at it. These factors need to be taken into account but won't be a focus of this study.

Due to a tablet being the primary hardware for digital interfaces in ambulances, this will be the focus of the design. Phones should be a strong consideration since they are also used in the ambulance but were out of scope for this project.

The AI input and output will primarily be based on trauma patients to create a focus of the project and make the interaction based on real scenarios, but the interface aims to work towards other scenarios as well. However, these will not be designed or tested.

The user test will only test the initial experience of using the interface. A real interface would be repeatedly used each day. How this daily interaction would impact user behaviour will not be investigated. The same applies to the long-term use of AI, and how this impacts trust and decision-making biases. Further, actual decisions based on AI will only be investigated on a superficial level, since it would likely require an unfeasible amount of users for significant results.

1.7 Research questions

RQ1: How can a dynamic AI-driven decision support be designed to align with the ambulance personnel workflow and work environment?

RQ2: How can a dynamic AI-driven decision support be designed to align with the structure of current IT systems used in ambulances?

RQ3: How important is explainability for ambulance AI-driven decision support?

RQ4: How can AI decision support be explained and visualised to increase trust and understanding of the system?

RQ5: How could AI-driven decision support impact ambulance workers' cognitive decision process?

2

Context

The following chapter goes through the context of the ambulance work environment. While the aim of the report is to apply to all ambulance systems, the focus is on Sweden generally and the Swedish region Västra Götaland(VGR) specifically, mostly due to its proximity to the authors, making it easier to study. Most of the contents of this chapter are based upon information gathered from the interviews with experts and ambulance personnel and relevant documents.

2.1 Ambulance organisation

Prehospital care is care that occurs outside of a hospital setting e.g. home, emergency medical service(EMS) vehicle, and incident hot spots. In the case of emergency ambulance care, providers include; EMS nurses or paramedics and emergency technicians (EMTs). The care provided in this setting is typically for patients who are experiencing an acute medical emergency or trauma, such as a heart attack, stroke, and vehicle accident. However, it can also be provided to less emergent cases only requiring transport or patients that appear to be healthy enough to stay at home.

The purpose of ambulance care, in the case of acute cases, is primarily to stabilise the patient's condition while providing rapid transport to a hospital for further treatment. Most of the time, there are two personnel in every ambulance, one driving and one taking care of the patient. In most cases, they both take part in medical decisions, although only one is fully responsible. The roles can be switched between cases, or based on who has the most medical expertise.

2.2 Ambulance workflow

The following workflow (Figure 2.1) is primarily based on our ethnography research conducted in Halland and VGR regions in Sweden.

Notification from SOS Alarm centre
The communication between SOS Alarm Centre and the EMS is done via the national RAKEL system; however, this is used by the police and fire departments. EMS personnel usually carry a telecom tool with this system, and every ambulance is equipped with one unit. Whenever a mission is sent out from the SOS Alarm Centre via the RAKEL system, this can be a prioritisation between 1-3 where a 1 is very

Figure 2.1: The flowchart represents EMS general workflow and routines.

acute and needs to be responded to as early as possible. Then the EMS personnel needs to confirm it; an estimated arrival time is reported back to the SOS Alarm centre, and that is when the mission begins.

Reach patient's location
The EMS personnel read through a brief description of the mission from the Emergency Medical Dispatch Centre (EMDC). If there is enough time, they can also view some of the patient's medical history through a laptop or a mobile phone in the ambulance. With this information, the EMS personnel may get a better view of the patient and prepare the appropriate equipment.

Assessment
When arriving at the patient's location, an initial assessment examines the most important vital signs, such as breathing and circulation. Suppose a life-threatening acute emergency has not been found, the EMS personnel collects information by interview and begins a physical exam for vital parameters (e.g. pulse and blood oxygen saturation).

In the Halland region, these parameters can be transferred wirelessly to the IT system, while in VGR, the parameters need to be written down on paper. Triage systems such as Rapid Emergency Triage and Treatment System (RETTS) or West Coast System for Triage (WEST) are used through a digital interface such as the one in Halland or a pen & paper system practice in VGR. An encyclopaedia of symptoms and treatment guidelines can help EMS personnel throughout the process. The parameters will be noted, and when the investigation is done, the EMS personnel converse about further actions in consultation with the patient, relative and or healthcare staff through telecommunication. The patient must approve the care decision when consent is possible, as the EMS nurses cannot force any treatment on a patient unless necessary.

Treatment
There are four primary choices of action; the patient could be treated on the scene, the patient should seek further care at a primary healthcare centre, the patient needs to go to a hospital but can wait for a taxi or hospital transportation to take them there or the patient needs to go to a hospital in an instant and is taken to the Emergency department (ED) by the ambulance.

Transfer patient
When the patient is taken to a hospital; an acute patient medical record must be completed beforehand and handed over to the emergency department in the hospital. Region Halland and some other regions (e.g. Växsjö, Kronoberg) utilise a digital record that can be transferred in real-time. In contrast, region VGR uses paper journals that must be handed over in person. Parameters, medical history, and the patient's status are noted. The EMS personnel who provide the treatments on the scene are noted as responsible for the medical decisions, such as triage and treatment.

Write a patient medical record journal
When the mission is completed after the patient is transferred to a hospital, the rule of thumb is the EMS personnel has 20 minutes to finalise writing the patient's

medical record in a web-based system, such as Ambulink. This takes place where it has access to a stationary desktop computer, often at the hospital or the ambulance station, but most ambulance vehicles are equipped with a laptop where it can be performed as well.

The ambulance typically receives a new assignment after 20 minutes. However, if there is a new assignment with the highest priority no. 1, it can be set directly after the previous one is completed. In these cases, the writing of the patient's medical record is postponed. In the worst-case scenario, the whole shift can go on without having enough time to transfer the notes on paper (VGR) to the patient's medical record. Thus, it has to be done at the end of the shift instead.

The regions like VGR allow access to patients' medical journal outcomes after visiting the hospital. This allows EMS personnel to check the results (optional).

2.3 Triage systems

The triage system is utilised as a tool for nurses to help them prioritise patients. Typically, the severity of cases is colour coded, with blue, green, yellow, orange or red. From yellow and upwards, patients should be sent to a hospital. In the case of red, the case is very urgent, with severe consequences if the patient does not receive healthcare in time. Orange is also urgent but less urgent than red. Yellow means the patient should be sent to the hospital but is not time-critical.

Several triage systems have been developed and established globally. However, the most common use of the triage system in Swedish healthcare is RETTS, according to Magnusson (2021). The system is based on vital parameters (VP), emergency systems and signs, and patient information such as age.

VGR uses its system, WEST (Tran, Winkler, 2021). It is based on the South African warning scale(SATS) (Figure 2.2), which categorises symptoms by severity, and guides triage based on a step-by-step method, first looking for the most severe symptoms and then going further. WEST also utilises the national early warning score (NEWS) (Figure 2.3) that continuously considers VP, giving a severity number based on how much lower or higher they are outside regular intervals. If one score is very severe, it can be enough to give high priority to a patient. Otherwise, a combination of scores is taken into account, and depending on how high the score is, the patient has ascribed a coloured prioritisation level.

2.4 AI-driven decision support for Trauma

Although the project is supposed to work with different AI models, for the sake of having a focus point, the project was based on an AI model for trauma prehospital triage, developed at Care@Distance by Anna Bakidou(2023). Similar models have been developed in the research group before such as an On Scene Injury Severity Prediction (OSISP) model for motor vehicle trauma (Candefjod et al., 2021). Bakidou's model is similar but used for other kinds of trauma as well, both blunt and

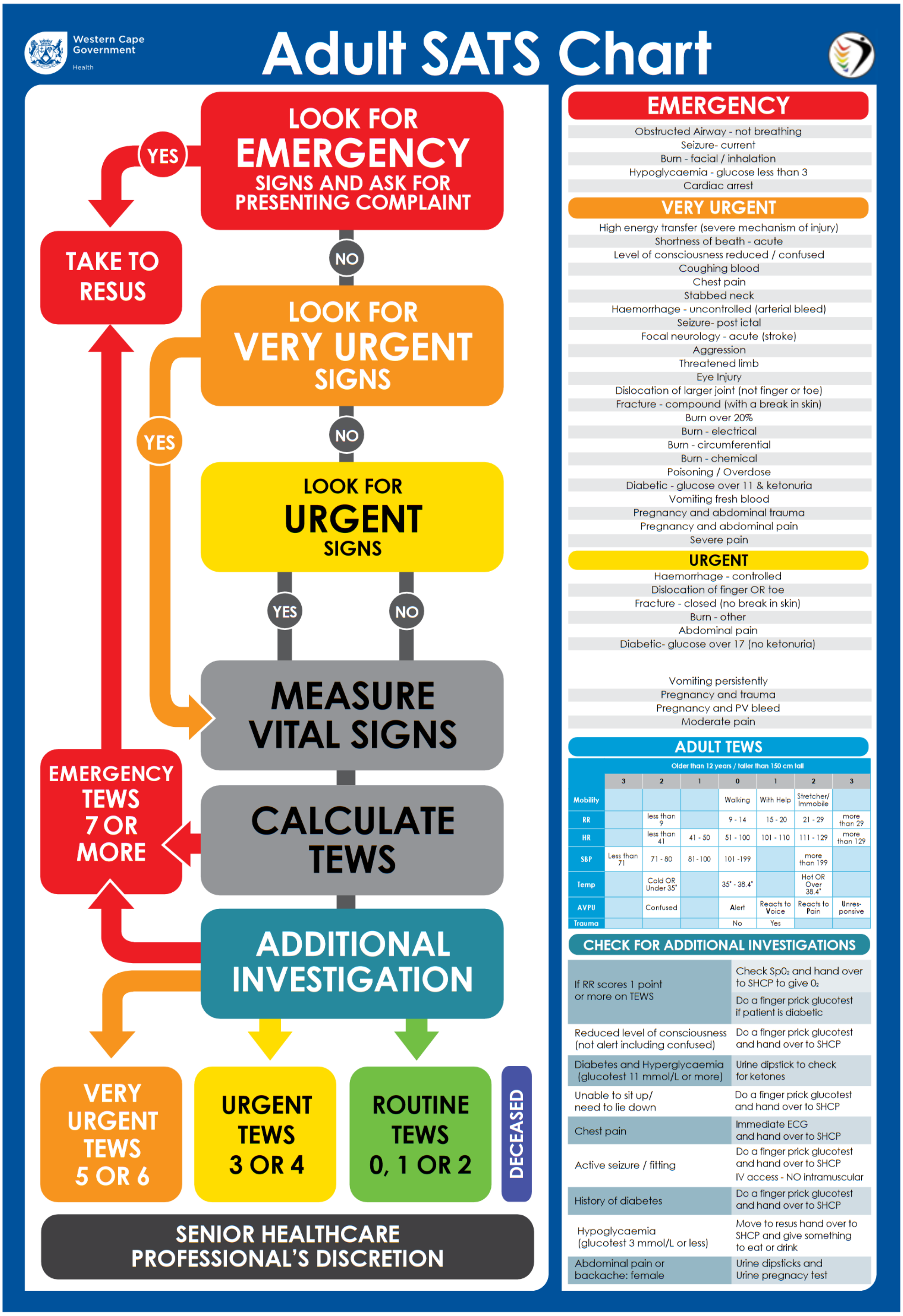
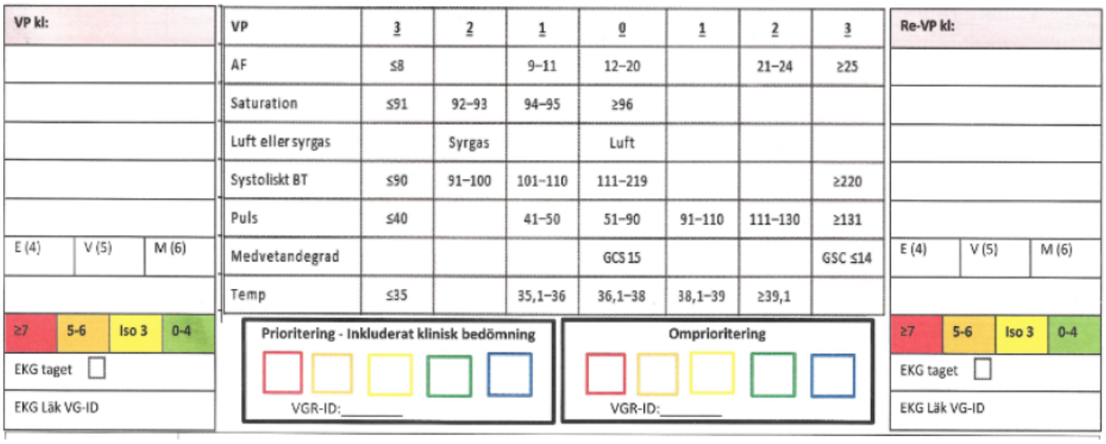


Figure 2.2: Adult SATS Chart, The South African Triage Scale (Cheema and Twomey, 2012)



VP	3	2	1	0	1	2	3
AF	≤8		9–11	12–20		21–24	≥25
Saturation	≤91	92–93	94–95	≥96			
Luft eller syrgas		Syrgas		Luft			
Systoliskt BT	≤90	91–100	101–110	111–219			≥220
Puls	≤40		41–50	51–90	91–110	111–130	≥131
Medvetandegrad				GCS 15			GSC ≤14
Temp	≤35		35,1–36	36,1–38	38,1–39	≥39,1	

VP kl:
E (4) V (5) M (6)
≥7 5-6 Iso 3 0-4
EKG taget
EKG Läk VG-ID

Prioritering - Inkluderat klinisk bedömning
VGR-ID:________

Omprioritering
VGR-ID:________

Re-VP kl:
E (4) V (5) M (6)
≥7 5-6 Iso 3 0-4
EKG taget
EKG Läk VG-ID

Figure 2.3: NEWS system from VGR Paper journal: Each parameter is given a score, they are then combined. If the score is seven or higher, the severity is red. 5-6, orange. 0-4, green. Yellow has separate preconditions.

penetrating, like falling injury, stab injury, explosion, etc. For training and validation, data from the Swedish national trauma registry (SweTrau) from between 2013-2020 with 47357 patient registrations after exclusion criteria are utilised.

The model uses a total of 13 different predictors, including body region, injury mechanism, intent of injury, age, etc. The combination of these predictors for each patient is then compared to the patient outcome during training. When validating, the model is granted a dataset that hasn't been trained on, is given the predictors for each case and makes its own prediction. Performance-wise, when set to the recommended level of over-triage at 25-35%, the model yielded under-triage at 8-25%. That is potentially a big improvement to current triage results, with current over-triage at 45.8% and under-triage at 40.4% according to the data in the SweTrau registry. This is also before the algorithms received any optimization for the trauma model which in this case only used the default settings.

Five different machine learning techniques were used to train the model, including more transparent techniques and those more like a black box such as an artificial neural network. However, both types delivered similar performance, meaning a transparent method could be used, which would allow for knowing more about what caused the AI to come to a conclusion, for example, which predictors were the most important. This is paramount in the case of healthcare, for allowing healthcare workers with understanding why the AI came to a specific conclusion.

3

Theoretical framework

The theoretical framework for this thesis formed a base for important design considerations and the general process of the project. This includes human factors, user experience design, design theory, healthcare regulations and guidelines, and ethical challenges concerning the use of artificial intelligence.

3.1 Human factors

3.1.1 Cognitive processes and bias

Wickens's attentional processing model(2021) shown in Figure 3.1, describes several information processing and decision-making stages. Stimulus creates sensory input through our sensory organs. This is then perceived based on knowledge stored in the long-term memory, which utilises working memory to process stored information, connecting it to the sensory input. Then, working memory is utilised to consider possible responses to the stimulus, finally leading to a response execution, which creates feedback with new information. Working memory is minimal, both in the number of units and time. Thus, throughout the design process, this will need to be considered to ensure ambulance workers only need to hold small sets of information at a time.

At every stage in the process model a pool of attentional resources is spent, meaning that if many resources are used in one stage, it can affect other stages, leading to errors (e.g., in perception or response selection) which could be detrimental to decide optimal care for a patient in an emergency situation. The project should aim for a design limiting the usage of attentional resources.

Ambulance personnel can often need to work with time-shared tasks, such as listening to information from a patient or monitoring vital signs while putting the data into a journal. According to Wickens (2008), the 4-D multiple resource model, Figure 3.2, predicts how different cognitive modalities can be used simultaneously while reducing the expenditure of resources from each other. (e.g., talking to a patient and reading instruction are both considered verbal/linguistic modalities). At the same time, the spatial task of putting on equipment and speaking are two different modalities and can share time-space. Thus, what modalities are used can be important considerations for the design.

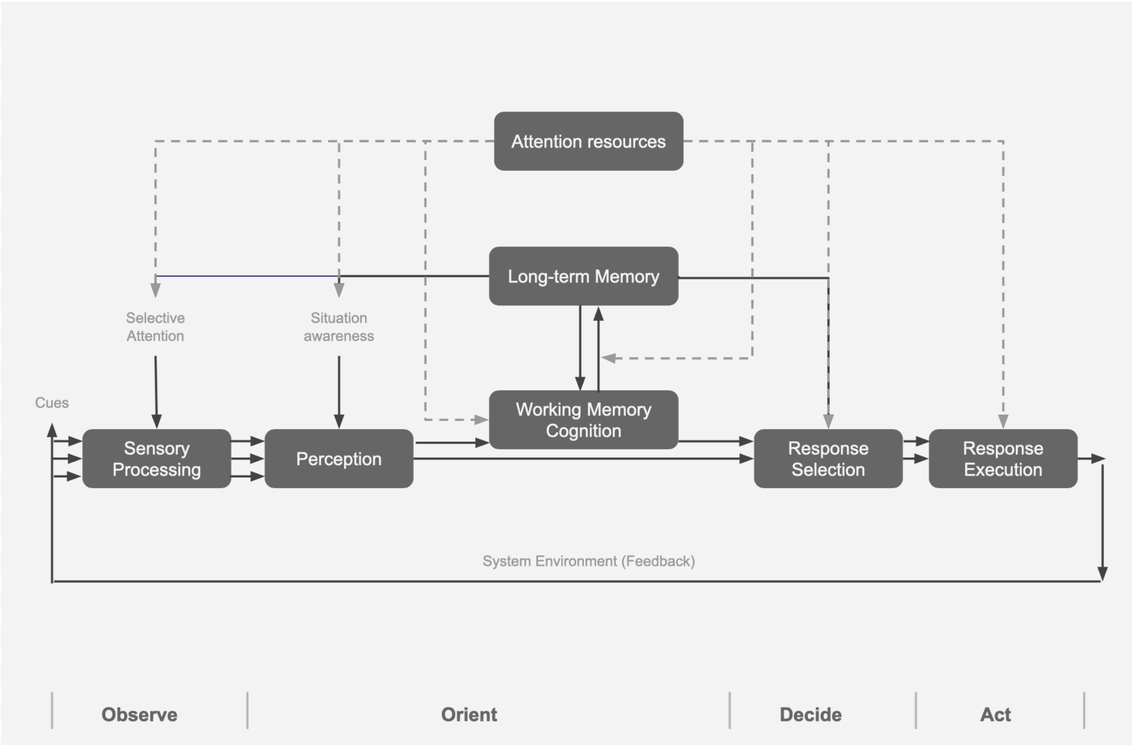


Figure 3.1: The human information processing model adapted from Wickens, Hollands, Banbury and Parasuraman (2016)

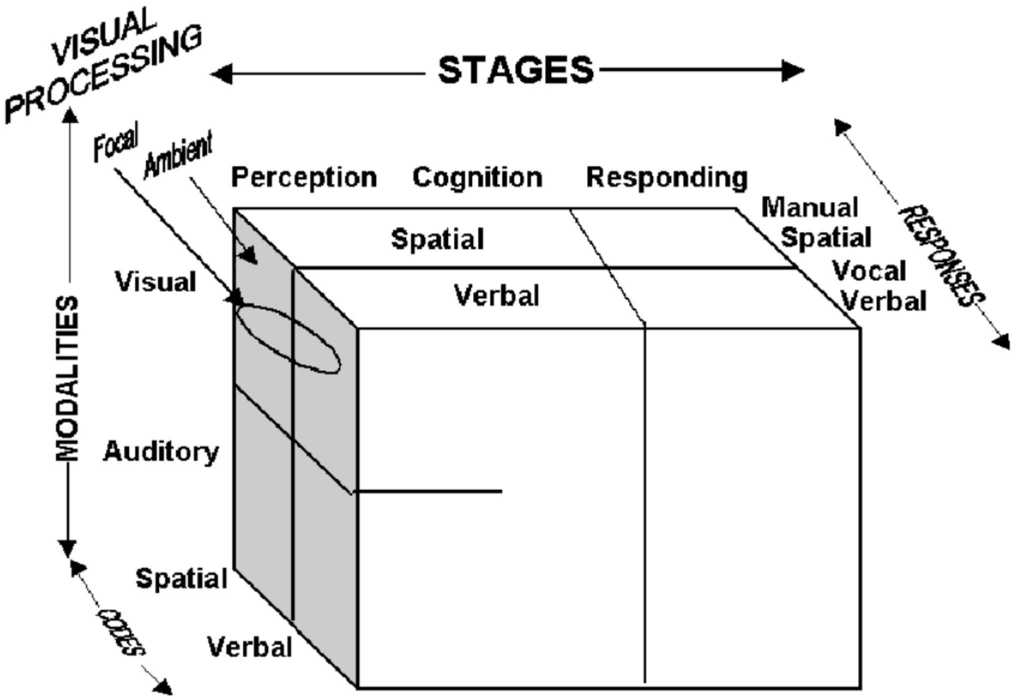


Figure 3.2: 4-D multiple resource model by Wickens et al.(2008)

A prominent decision-making theory is that we have two modes of thinking, system 1 and system 2, seen in Figure 3.3 (Kahneman, 2011). System 1 is quick and automatic, taking mental shortcuts called heuristics in order to make conclusions. System 2 is analytical, slow, and requires more energy and concentration. Most decisions are of system 1, and in a situation such as an ambulance where stress, fatigue and time limits can be significant factors, system 1 is a likely major factor. This can lead to a more decisive influence of biases and incorrect heuristics. This includes biases such as anchoring, where the first piece of information influences the rest, meaning when and how to present information in an IT system can lead to different consequences. Another is the framing effect, where talking about the risk of death has a different effect than the chance of survival, although the numbers could be the same. Investigating the most apparent biases in prehospital care and using the design to limit them could improve decision-making and reduce errors.

Figure 3.3: System 1 and System 2, adapted from Kahneman (2011).

3.1.2 Gestalt principles

The gestalt principles describe how we perceive group objects (Todorovic, 2008). For example, the proximity principle tells us that objects close together are within the same group. Similarity says the same for objects that are similar to each other, and the common region principle says the same for objects that share the same closed region. See Figure 3.4.

An interface in the medical field can include several kinds of information, such as vital parameters, personal information, and guidelines. To make sure these are not accidentally grouped together as the same objects or objects of the same importance, following Gestalt principles are necessary.

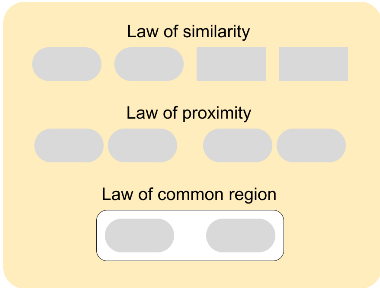


Figure 3.4: Three examples of Gestalt principles, adapted from Todorovic (2008).

3.2 User Experience Design

User experience (UX) design is acknowledged as a framework for understanding the interaction between users and artefacts, see Figure 3.5. UX design typically focuses on creating compelling and satisfying experiences for them (Norman & Nielsen, 2010). User experience can be defined as "a person's perceptions and responses that result from the use or anticipated use of a product, system or service" (ISO 9241-210). According to Jesse James Garrett (2010), the term 'user experience' is generally applied to the positive, neutral, and negative emotions felt whilst interacting with computer systems and user interfaces and is equally applicable to any other instance where a human uses a product, object, or service. He stated, "Every product used by someone has a user experience: newspapers, ketchup bottles, reclining armchairs, cardigan sweaters".

The UX design relies on various principles, concepts and methods and is set on various disciplines such as psychology, sociology, and human-computer interaction (HCI) (e.g., Figure 2.). UX design emphasises the importance of understanding user needs and goals and designing products that meet their needs seamlessly and intuitively. UX design provides a powerful framework for creating digital products that are both effective and enjoyable for users (Garrett, 2010). The process of UX design practice will involve a deep understanding of user behaviour, preferences and pain point issues, including the broader social and cultural context in which the product will be used. The process generally interacts and continues improving products, systems or services. Designers must be willing to test and refine their designs based on user feedback and to adapt to the change in user needs and market trends. This would encourage designers to experiment, try new approaches, and be open to feedback and criticism.

3.3 Double diamond model

The activities and methods will be performed for this project following the double diamond model, see Figure 3.6. It is a process introduced in 2004 and has become a well-known model for designers and non-designers to systematically engage in problem-solving (British design council, 2019).

Regarding this project, developing a future IT system is a complex and dynamic

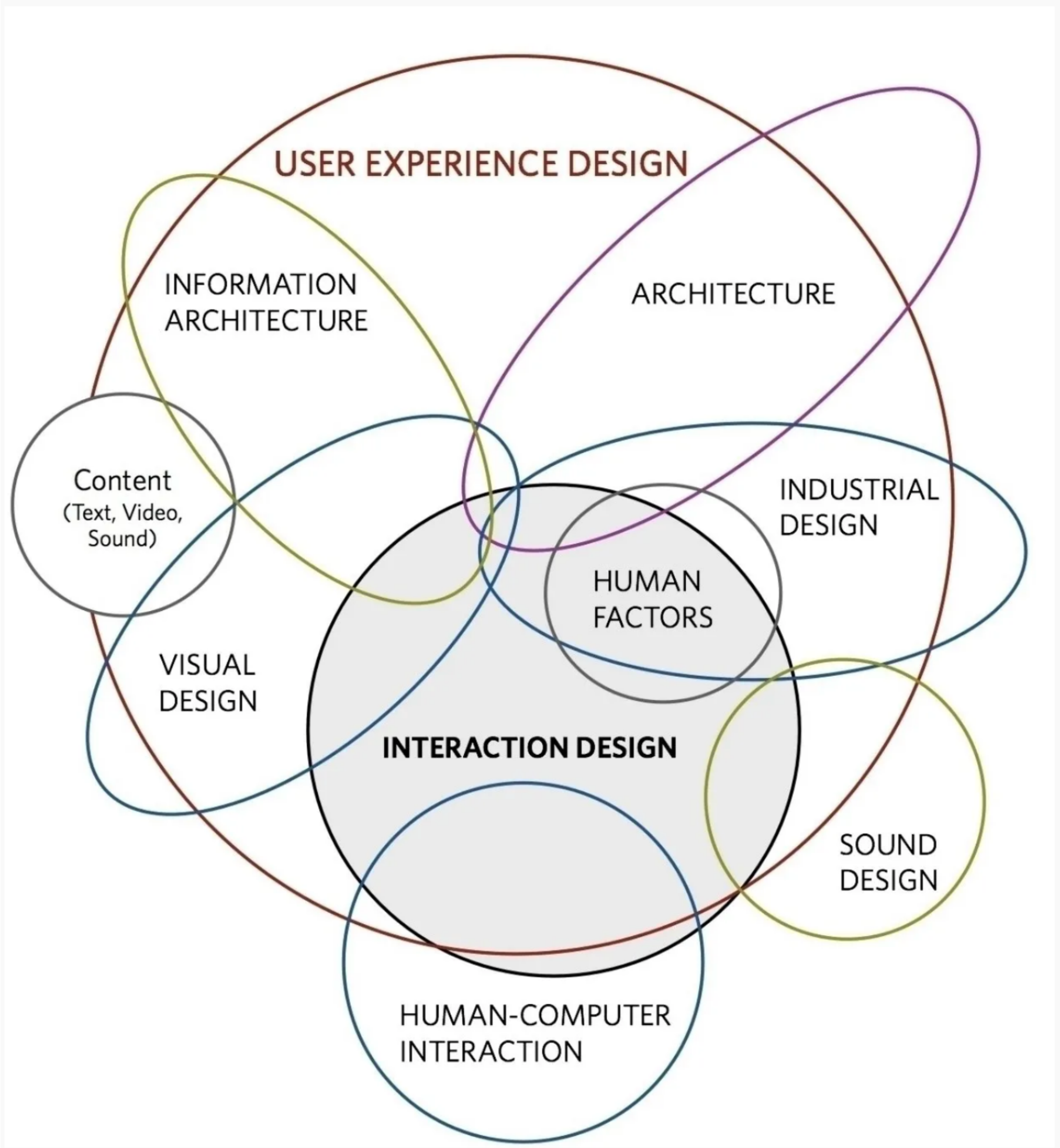


Figure 3.5: Disciplinary map of user experience design to other design fields (Interaction Design Foundation, 2002).

project requiring an understanding and effective approach. Therefore, the double diamond model is a framework that can offer several advantages. First, the divergent part of the model allows for the exploration of these complex tasks related to the work environment, current decision support tools, diverse technical solutions and organisations behind such solutions. When such findings have been made, the project can develop into a narrow scope, a comprehensive system to improve patient outcomes that meet the user's needs and stakeholder's expectations.

However, the limitation of the double diamond model that Sharp et al. (2019) mentioned is; It can be time and resource intensive, relying heavily on user feedback which may be limited in some cases. The model may only sometimes fully address broader social and cultural factors, and its flexible approach may need more explicit direction. It was crucial to consider the potential limitations of the double diamond model for this project and utilise it based on the context and goals.

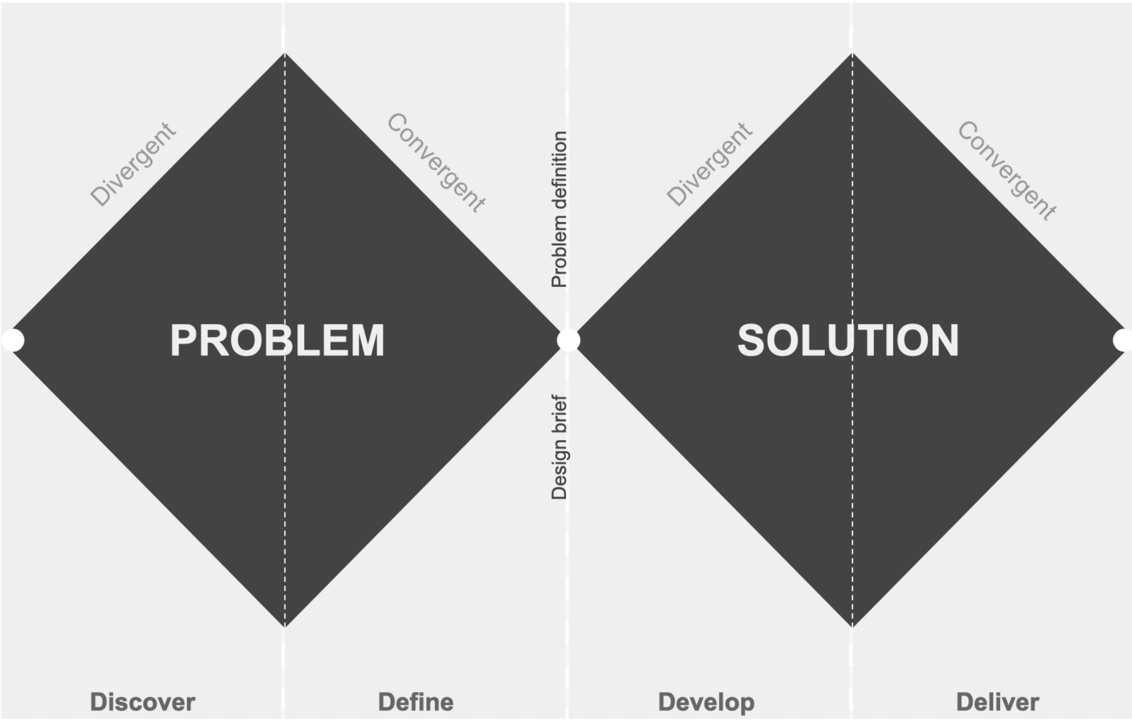


Figure 3.6: The double diamond of design, adapted from Sharp et al. (2019)

3.4 Swedish healthcare regulations, guidelines and ethics

The project will focus on the Swedish EMS system. While not necessary for this research to succeed, if the project is to be able to present a feasible solution according to Swedish standards, it will need to try to follow Swedish regulations. Swedish healthcare is governed by Hälso- och Sjukvårdslagen (HSL) [The Health and medical care act], decided by the Swedish parliament; Chapter 5, 1 § states that all healthcare shall be, as translated by the author of this paper:

1. Be good quality with excellent hygienic standards.
2. Satisfy the patient's need for security, continuity and safety.
3. Build on respect for the patient's self-determination and integrity.
4. Promote good contact between the patient and the healthcare staff.
5. Be easily accessible.

In particular, points 2-4 are relevant since they can inform a focus on the patients' needs in the design. While the project will focus on the user/paramedics, the results should promote the patients' needs to be stated in HSL. An important consideration could be how to design to keep respecting the self-determination and autonomy of a patient while an AI makes a recommendation that might be more difficult for humans to interpret. In addition to HSL, ambulance-oriented healthcare has its own set of regulations, as translated by the author of this paper (SOSFS 2009:10, chapter 2, 1 §);

- Information cannot be changed by mistake, unauthorised use or due to malfunction or any other disturbance.
- Every incoming and outgoing alarm, as well as other communication via logs, can be traced to operators at the alarm centre as well as health and medical personnel inside or outside an ambulance.
- The requirements of regulation, EU 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free flow of such data and the repeal of Directive 95/46/EC (general Data Protection Regulation) as well as the requirements in the law (2018:218) with supplementary provisions to the EU's data protection regulation and in the patient data act (2008:355) are met (HSLF-FS 2018:38).

It will be impossible for this project without judicial expertise to abide by all regulations, but they can still act as guidelines for the design process. In particular, the first paragraph here highlights the importance of the design to decrease the risk of user error, either through prevention or easy correction.

When deciding which patients to help, healthcare workers are regulated to follow the prioritisation platform with three guiding principles(Prop. 1996/97:60), as translated by the author of this paper:

The human value principle is that all humans are equal in value and have the same rights independent of personal attributes and roles in society, Needs-Solidarity principle; resources should be distributed depending on the need, Cost principle when choosing between different operations or measures, a reasonable relationship between costs and effectiveness, measured in improved health and increased quality of life, should be sought.

The design needs to support these principles because paramedics constantly have to make decisions and prioritizations during the triage process. It also allows for a better understanding of which principles guide EMS decision-making during the empathising part of the project.

3.5 Ethical challenges

Healthcare is a sector where consideration of ethics is central. While doing observations of ambulance workers, there is a need to ensure to not interrupt their work process in any way since this could cause risks for them and their patients. Situations can be witnessed where respect for the patient's privacy is paramount and information that can be linked to them can not be spread under any circumstance.

Carefully considering the design's effects on healthcare workers' working environment was required throughout the design process, as well as patient safety and privacy. For instance, some designs might give an AI algorithm more data, allowing it to make better decisions in the future. Still, it is possible that the interaction would be too slow, risking patient safety and requiring information that might go against the ethics of patient privacy. The design would be used continually every work day, so considering mental load and stress factors is paramount for the well-being of ambulance workers.

3.6 Artificial intelligence

3.6.1 Artificial Intelligence (AI)

Russell& Norvig (2010) describe in the book that AI as a field of study aims to develop intelligent machines capable of performing tasks that typically require human intelligence. AI covers various approaches and techniques, (e.g. machine learning, natural language processing, and robotics). The stage of AI is based on the idea of creating algorithms and system models that can process large amounts of data and learn from collected data to make predictions or decisions. Russell and Norvig (ibid.) also mention that unsupervised machine learning involves developing algorithms capable of learning from data without being explicitly programmed. This allows algorithms to improve their performance over time without needing direct input from humans. A rising number of AI developments are also leading to ethical discussions and social issues, such as the potential for bias and discrimination (Zuiderveen Borgesius, 2018).

3.6.2 AI in healthcare

AI has played a significant part in the public consciousness, gaining enormous publicity through models, for instance, chatbot ChatGPT and picture generators, e.g. Dall-E 2 and Midjourney. AI has become a more natural part of ordinary life through different tools and natural language processing with speech dictation. In addition to general use cases, specific AI tools have great potential to be used in professional fields like healthcare. Much research has been done on AI image classification (Razzak et al., 2018), such as detecting diabetic retinopathy (Abràmoff et al., 2018), cancer diagnosis and prediction (Goldenberg et al., 2019) and Covid-19 diagnosis(Hassan et al., 2022). Overall, AI can potentially improve diagnosis and prognosis compared to healthcare professionals, allowing for improved decision-making.

In healthcare sector has specific requirements for AI healthcare. For example, WHO (2021) gives guidelines for using AI development, including transparency, explainability and the need to work against bias. However, AI can often be described as a black box, where we do not have the tools to understand too complex algorithms. Thus, AI needs to be developed with this in mind because it is essential for both developments of AI, patient autonomy and safety. Furthermore, it requires explainability; thus, medical professionals must make informed decisions. Finally, AI is often trained on data that already are formed from human bias, which needs to be heavily considered to reduce significant errors in AI decision support.

Emergency services (EMS) play a pivotal role in healthcare outcomes and introduce new challenges compared to the AI tools previously mentioned. For example, an ambulance rarely has the same tools for diagnosis as a hospital, there is a wide variety of patient symptoms, and time can be minimal with acute symptoms. AI can potentially help triage patients, both in speed and in the case of reducing over-triage and under-triage (Buendia et al., 2015; Keselman et al., 2022), increasing patient safety and reducing resource depletion.

AI models require real-time digital data to function, and while many current IT systems have the potential to provide it, they often need to be appropriately used in practice (Porter et al., 2020). In addition, they can be suboptimal for the complex workflow of EMS personnel, resulting in the input being done after the patient has already been sent to a hospital. A design that considers the many factors impacting EMS and looks into cases where IT implementation has been successful is thus paramount.

3.6.3 Interaction and Artificial Intelligence (AI)

Interaction design and AI are two things that are increasingly intersecting or often come across each other. Interaction design focuses on designing effective, intuitive, and feasible products, systems and services for users. In contrast, AI involves developing intelligent machines capable of performing tasks that require human intelligence. As AI becomes more advanced and more involved in direct human interaction, it is used to enhance and improve the user experience in digital products and services. For instance, AI-powered chatbots can provide personalised and responsive customer service. At the same time, machine learning algorithms are used to personalise content, recommendations and search results.

However, He et al., (2022) discussed designing effective and engaging interactions between humans and AI-powered systems may encounter several challenges, such as designing transparency and understandable interactions with users so that they can trust and feel in control of the system, including designing inclusive and equitable interactions that prevent biases or discrimination. To address these challenges, designers might consider AI's ethical and social implications in their design process and develop user-centred, inclusive and transparent interactions.

4
Methods

This chapter goes through the methods used during the study. It provides an overview of the methods themselves and why they are used for this project. Some of the methods had minor modifications to be better adapted to the context of the study.

4.1 Literature study

A literature study provides an overview of the current state of knowledge and identifies gaps or areas for further research. The literature study will include an extensive analysis of research articles, books, reports, and other sources of information related to the project. Considering this, conducting a literature study would be in several stages. The first stage is identifying a research question or topic that is specific and narrow enough to allow for a comprehensive search of literature but also broad enough to capture the key concepts and issues related to the topic

The next step would involve searching for relevant literature using databases, online libraries, and other sources of information. The search terms should be carefully selected to capture all relevant literature. After identifying potentially relevant literature, the quality and relevance of each source need to be filtered and evaluated. The evaluation process included reading each source carefully, assessing each study's methodology, research design, data analysis and conclusions and determining its relevance to the research questions. Finally, the findings should include synthesised analysis, summarising each study's key findings, identifying patterns and themes in the literature, and drawing conclusions (Martin et al., 2012).

4.2 Ethnography research

Ethnography is a qualitative method to study people and their culture in their natural settings, typically over an extended period. Hammersley and Atkinson (2007) explain that this method aims to provide a comprehensive understanding of a culture or subculture and its beliefs, practices, values and social norms. Ethnography research involves prolonged engagement and observation, including conducting interviews, taking notes, collecting artefacts, and participating in their cultural environments.

In this case, ethnography research was done in two ways: following ambulance workers for a day in Halland and visiting the primary ambulance central in Gothenburg.

Halland uses a digital interface, and Gothenburg uses pen and paper; thus, visiting both regions was to gain a more comprehensive view of the field. In addition, both were important to understand the broader context of the ambulance work environment. This includes gaining insight into what tools are used, how personnel interact with patients and relatives, how they react to emergencies, work culture, and interaction with the IT-/pen & paper system.

4.3 Interview

Interviews are a great way to empathise with the users. Therefore an interview is an excellent method to gain an in-depth understanding of users' values, perceptions and experiences. The method allows us to ask specific questions while remaining open to exploring the participants' points of view. The interview methods are flexible and often combine with other research methods, such as usability tests or surveys, to gain better insights into objective results by asking users about them and to elicit their subjective opinions on the products or interactions (Martin et al., 2012).

To avoid bias, during this project's exploration phase, the context expects to be a busy environment. For better results, The interview shall be combined with the ethnography method in terms of contextual inquiry. Contextual inquiry can be made in different manners. For instance, the designers ask participants questions simultaneously while observing their interaction with a product or while they are performing a daily activity. The user is also asked to explain the interaction as though explaining it to novice users. Another way is to observe first and refrain from asking the participant any question until afterwards to avoid influencing or interrupting their behaviour.

Interaction design foundation (2002) has addressed the limitations of this method. "User interviews can be very informative and helpful, but only if they are used correctly and for the right things. Therefore, it's important to know what you can expect to get out of interviews and what you shouldn't expect to get out of interviews." "What users say and what they do are different," said Jakob Nielsen, Usability Expert and co-founder of Nielsen Norman Group.

4.4 Personas

Personas are a method to consolidate archetypal descriptions of user behaviour patterns into visual profiles that aim to humanise design focus to use in test scenarios and design communication. Personas typically present in short profile descriptions include a name, a photograph or chest, and a narrative story describing in detail such as living situations, goals, and behaviours relevant to the design question.

Personas are used as human references by the team throughout the design. Often helpful in all project phases, e.g., developing, discussing, and presenting product or system design in the definition and ideation phase. They are also valuable for checking use scenarios and highlighting positive experiences and potential breakpoints. Personas provide a persuasive human reference when communicating research summaries and scenarios to clients. (Hanington & Martin,2012)

4.5 Sketch

Sketching is an important activity in the design process, as it facilitates designers in exploring and conveying ideas visually. Crazy 8's and Solution Sketch are two sketching exercises frequently employed in the ideation phase of the design process to generate and refine design concepts. Crazy 8's is a high-velocity sketching technique- and Solution Sketch focuses on iterating and refining a single design concept. Crazy 8 involves folding a sheet of paper into eight equal sections and setting a timer for eight minutes. Designers sketch as many ideas as possible within the timeframe, with the goal of generating at least eight unique concepts. This technique's rapid and voluminous idea generation stimulates creativity and helps designers overcome design blockages (Reynolds, 2018).

Once a set of ideas has been generated through Crazy 8's, Solution Sketch can be used to develop a specific concept further. Solution Sketch is a method that involves detailed sketching of a single design concept while also considering its feasibility and implementation. The primary objective of Solution Sketch is to create a well-defined and elaborated design concept that can be presented to stakeholders for their feedback (Meroni et al., 2019). The combination of Crazy 8's and Solution Sketch provides a robust ideation and design iteration framework. Crazy 8's fosters designers to generate many ideas rapidly, while Solution Sketch provides a structure for refining and developing a specific design concept. Together, these techniques can assist designers in overcoming creative barriers, producing innovative solutions, and refining design concepts into practical solutions (Cooper et al., 2014).

4.6 Prototyping

Typically, prototyping involves creating an initial version of a product or feature that allows people to interact with it. The purpose of a prototype is to convey abstract concepts such as features, functionality, potential benefits, risks, and implementation costs. Then, testing prototypes with real users provides insights into what works, what does not, and whether it is the right solution to pursue. It also allows continuous iteration without spending needless resources on something that will not work satisfactorily.

A prototype is great for gathering feedback to keep improving the design. However, prototypes can vary in fidelity, and the quickest and most cost-effective ones are typically based on initial wireframes. Wireframes are in grayscale layouts that capture the structure, navigation, information architecture, layout details, and interfaces. Using sketches, task flows, and sitemaps to consolidate this information in wireframes is often an intermediate step between sketching and prototyping. However, depending on the design stage and prototype objectives, a prototype can be created by rapid paper prototypes, clickable wireframes, hand-coded interfaces, or high-fidelity prototypes resembling the final product. As a general guideline, it is beneficial to prototype early and frequently. Then, most importantly, test the prototypes and iterate based on feedback and insights from these tests (Houde and Hill, 1997).

4.7 Usability test

Usability testing is commonly used in design practice to evaluate systems by testing with end users or representative users. However, there are different ways of conducting usability testing, such as formal, informal and pilot ("down the hallway") tests.

- A formal usability test is a traditional way. Typically the researcher invites participants to the lab and performs the test or testing in their work environment to get a closer result to the end product.
- Informal testing is usually minor testing that can be done over time, with one or two users per week.
- The "down the hallway" test is "quick and dirty" testing similar to a pilot test that involves friends or colleagues who are testing the designs. Usually, this type of test can lead to bias; therefore, it is helpful before properly testing a design on behalf of representative users.

The usability test includes creating a plan to scoop out the test. The plan consists of user tasks and the sample issue, which will be given to the user to solve; recruitment criteria, scripts and discussion questions, and logistical information followed by facilitating or moderating the test. Then we can analyse data and finalise the test report to communicate the findings. Documenting the result is essential to help other people who were not involved in the testing and is a reminder for researchers when they need to revisit it (Rubin & Chisnell, 2008).

Interaction design foundation (2002) also describes usability tests as structured research sessions that help us understand apparent problems with a particular design. They are a trial run for a change to our product where we can see if potential users are likely to need help to complete specific tasks. Such a test usually involves observing users attempting to complete tasks and can be done for different designs. It is often conducted repeatedly, from early development until a product's release.

In this project, we have combined usability testing with the Think-aloud protocol. The think-aloud protocol is a method that requires participants to verbalise what they are doing and thinking as they complete a task, revealing aspects of an interface that delight, confuse, or frustrate (Hanington & Martin, 2012).

5

Results

This chapter starts with the results from background research. It then moves on to the design process, followed by the final design and its evaluation.

5.1 Literature study

During the divergent phase of the project, a literature study was conducted. The main aim of the literature study was to gain insights into the current state of IT systems used in ambulances, the performance of current triage systems, as well as current uses of AI in healthcare in general, and explainable AI(XAI) in particular. The secondary aim was to find gaps in current research. The main sources of searches were Google Scholar, Pubmed and Chalmers Library. Also, some articles were added as per recommendations from experts, supervisors or stakeholders. Keywords included, but were not limited to: "AI", "AI healthcare", "explainable AI", "IT-support ambulance", "triage models", "over-triage" and "under-triage".

In addition to work directly related to the field, literature studies were used throughout the project to gain insights into relevant work within user interface (UI) and human factors to help guide the design process. Many findings and articles are used throughout the text, but some key findings are summarised below.

5.1.1 Current use of it-support

Current it-systems used in ambulances in the UK are often underwhelming, leading to workers noting information on their hands or pen and paper before putting it into the system only after the patient has been sent to the hospital (Porter et al., 2020). Thus the information can't be used for digital decision support in an early stage. Key issues were that many it-systems were still in a transitory phase and that there were few standards for hardware and software, providing challenges for interoperability. Providing the correct information to the correct medical professionals was also a challenge.

5.1.2 Current triage decision support in Sweden

Magnusson (2021) compares two dominant triage systems in Swedish ambulance care, the Rapid Emergency Triage and Treatment System(RETTS) and the National Early

Warning Score (NEWS) (Royal College of Physicians, 2017), showing that RETTS has higher sensitivity for time-critical patients but less specificity than NEWS. Both systems have issues with both over-triage and under-triage, with about one-third of children being either over- or under-triaged and under-triage in general being an issue when evaluating the elderly.

5.1.3 Status of Explainable AI in Healthcare

While several applications for AI were found, described more in the background section, research on explainable AI (XAI) was very limited. Antoniadi, et.al., (2021) reviews the current literature and finds that it is only in very recent years that XAI in healthcare has really started to be explored. The majority of XAI was used for tabular data while research on XAI used for text analysis was the least common. XAI can improve decision-making and increase trust in the system. However, XAI research in clinical decision support was extremely limited, especially when it came to user research.

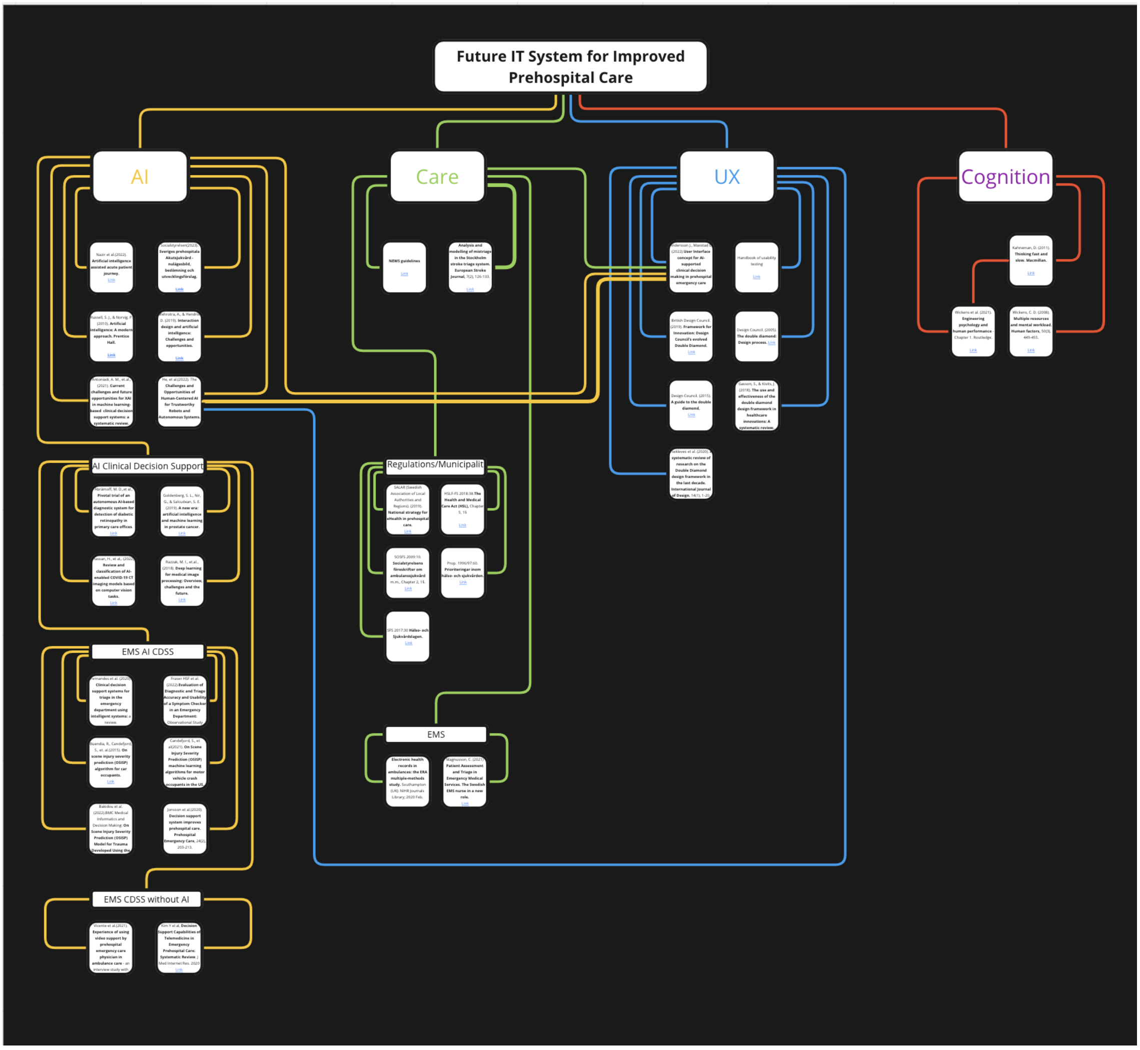


Figure 5.1: Gap analysis map of this project's literature review

5.2 Ethnography research

This session will discuss ethnography research on how we used the methods adapted to the project in best practice. And the results were interpreted from this particular field study. However, in a phase, we aim to understand the user's point of view and emphasise as much as possible their role to be able to learn their daily routine.

This project got an opportunity to explore two different regions in Sweden. One is Halland, and the other is VGR. Furthermore, due to the positive response from the Halland region, the team had the opportunity to follow up their routine for a day from 7.00 o'clock to 20.00 o'clock (one full shift).

5.2.1 Empathy

The primary approaches used in ethnography research are observation and contextual Inquiry. The observation is typically "learning by watching" while taking notes on the essential things happening that are considered valuable later on. The most important thing is to act calmly, dress like them, be neutral as much as possible, try to understand what is happening and interfere less with the EMS work routines, in combination with the form of an interview (contextual inquiry) that is described as "asking while doing it". If they are at some point wondering about their action or confused in some ways, we rather question "the right way or can ask later on as appropriate" (Hanington & Martin,2012).

Figure 5.2 shows the results from EMS's general context in the ambulance workflow and work environment interim of understanding and empathy. Unfortunately, some images are not allowed to be shared in this report. There were several actions and tools we didn't understand at first, but everything was noted after receiving answers from them.

5.2.2 Strategy and product objective

Figure 5.3 displays an overview of the existing system on the market. It compares clear user Interfaces from different products used in emergency service. This overview provides additional insights beforehand that can be considered as State-of-the-Art (SOTA). Oxford Dictionary (n.d.) describes SOTA as "the most recent stage in the development of a product, incorporating the newest technology, ideas, and features."

SOTA's input has motivated us to learn the design patterns and the current trends of the existing system. It allowed us to analyse the pros and cons of the popular systems currently on the market that may lead to a better design later on. Learning their strategy also notified us about the most obvious errors and weaknesses that might be prioritised in the next system update.

5.2.3 Identify and understand user

To be able to get a better comprehension of the user and their workflow, we have created personas to consolidate archetypal descriptions of user behaviour patterns,

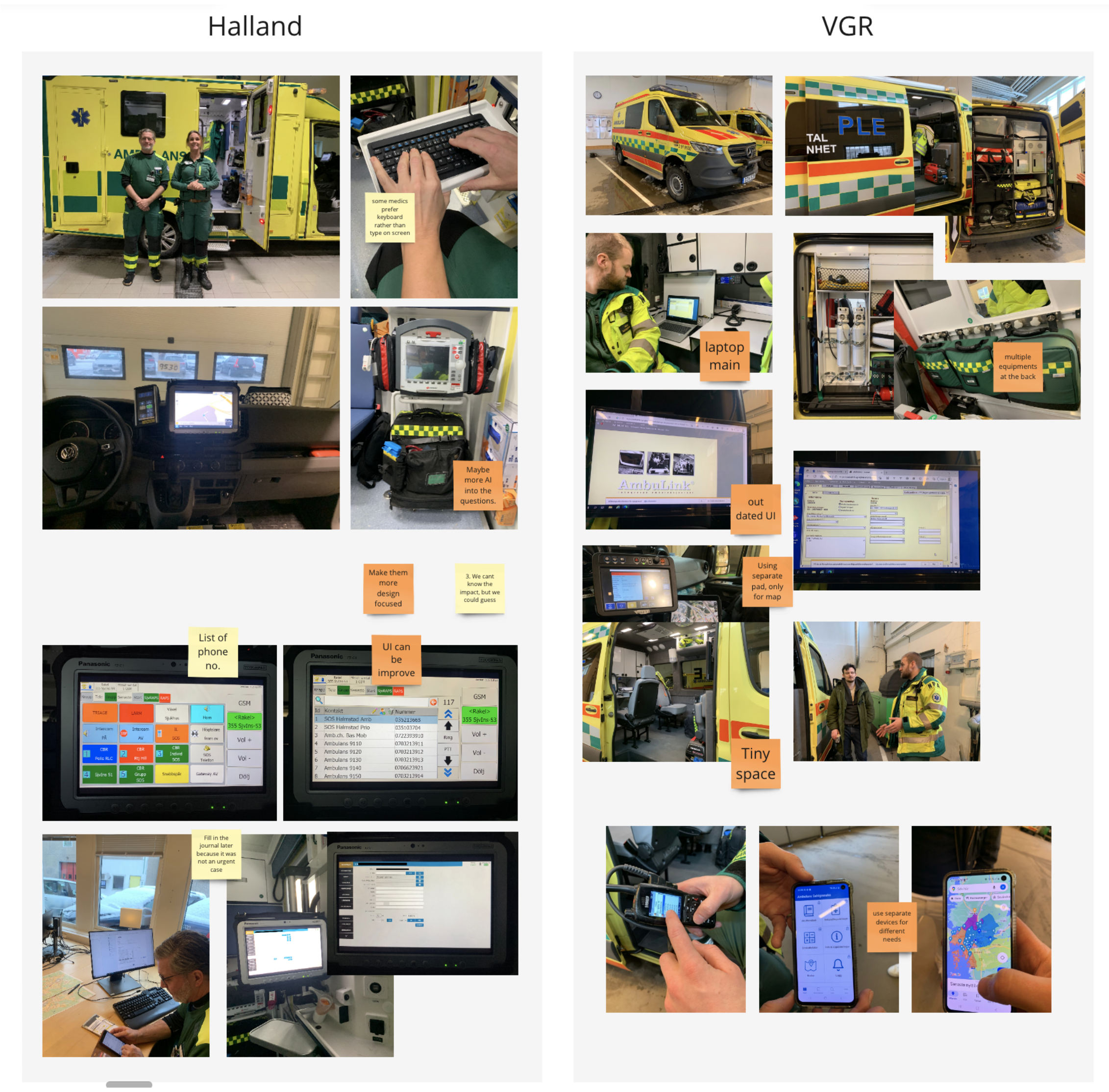


Figure 5.2: Ethnography research; Observation & contextual inquiry from regions, Halland and VGR)

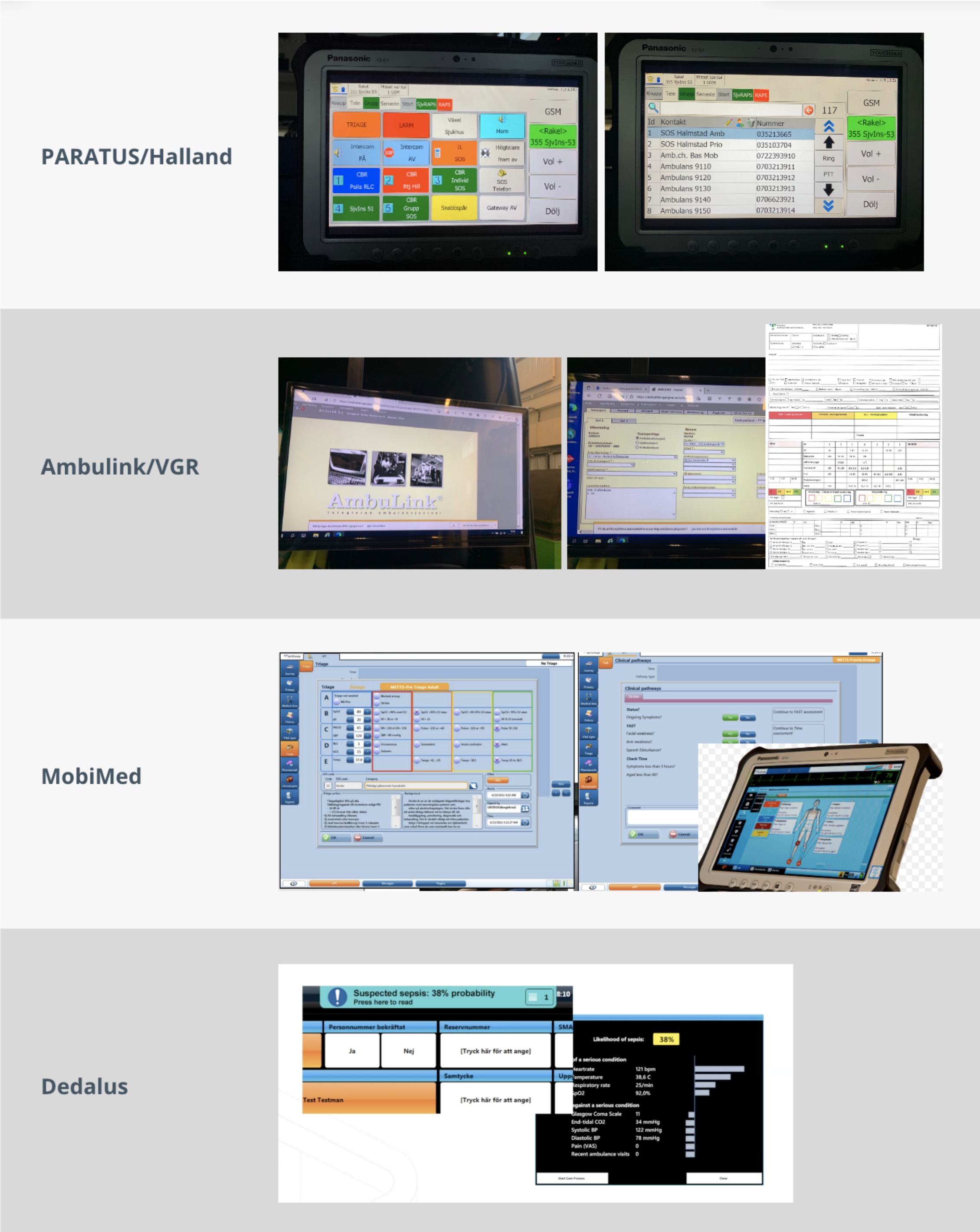


Figure 5.3: SOTA table showing product diversity.

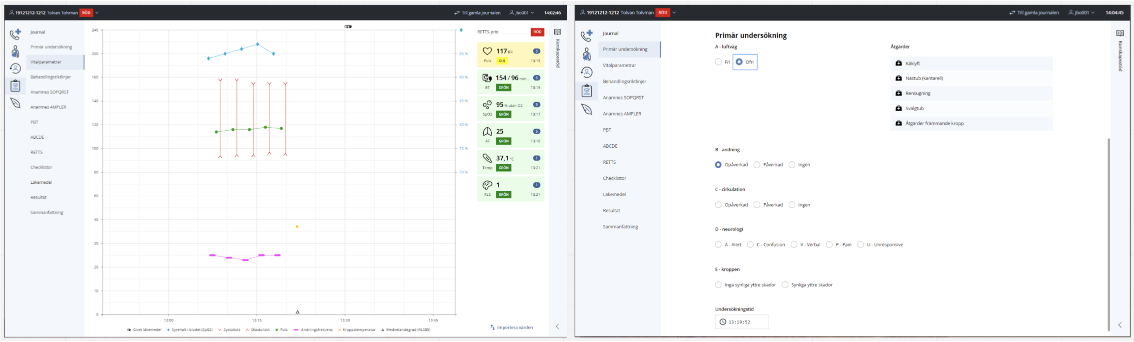


Figure 5.4: The second latest version from Paratus.

see Figure 5.5, and a journey map, see Figure 5.6, that analyses the context based on the observation data. The journey map is meant to provide a heuristic analysis narratively for the team. For example, it may begin with a brief explanation of users and behaviours in their work environment setting, followed by phases, actions, trends, and narrative facts.

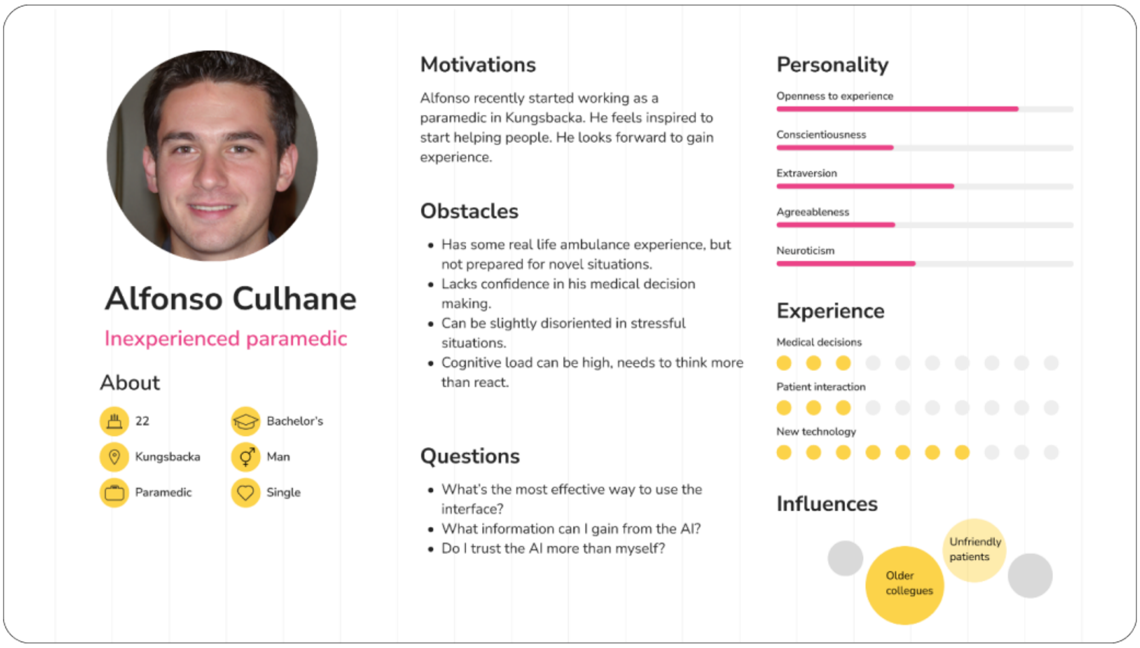


Figure 5.5: One out of four developed personas. Fictional profiles that represent user behaviour patterns based on collected data.

As a result, the journey map should review a clear understanding of user touch points, a clear understanding of the channels in which actions occur, an understanding of any other actors who might alter the user experience, and a timescale.

In conclusion, the journey map presents a heuristic view of the observation addressing problems and identifying activities, action, interaction, and satisfaction in the EMS workflow based on the user personas, similar to the AEIOU framework which stands for activities, environments, interactions, objects and users. Hanington & Martin (2012) describe this framework as an organisational framework reminding the

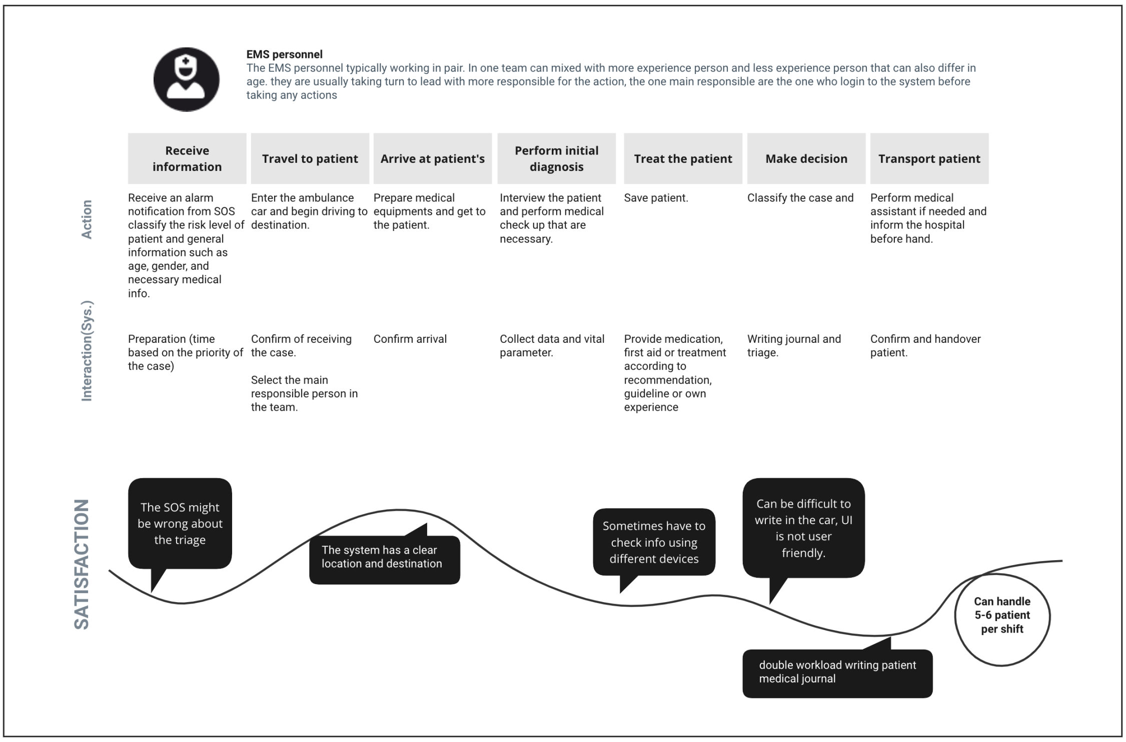


Figure 5.6: Journey Map

researcher to attend—document, and code information under the guiding taxonomy the name is based on.

5.3 Expert interview

Stefan Jönsson - Future of healthcare in VGR
Highlights:

- New IT-solution in VGR called Millenium on the path to be implemented
- Live data transfer with high interoperability
- Ambulance part of Millenium planned but hasn't started proper development

An interview was conducted with Stefan Jönsson, process manager of emergency care, working with regional development at VGR. He has responsibilities for the implementation of a new IT system for ambulances and emergency departments. Currently, different hospitals in VGR use different journaling systems, making it difficult or impossible to transfer data between them. A new system called Millenium is currently in the process of being implemented, allowing for higher degrees of interoperability. The planned launch date is 2025 but without all functionalities.

The system for ambulances hasn't started development yet, however, some functionalities already existing in Millenium might be used for it. The plan is to have a joint journaling interface, both for the ambulance and the ED. Both should be user-friendly and quick to use. The journaling works in real-time, with different healthcare systems being able to add information to a patient. For example, the emergency dispatch centre might already have information on the patient- such as the type of accident and that could already be filled into the same system used by the ambulance. When passing through the healthcare chain, the patient should be continuously evaluated. Access should be granted to previous journals within the system, combining data from different sources such as the 1177 medical advisory call centre, local hospitals or specialist hospitals. Finally, since the EMS IT system will be integrated with the hospital system, ambulance personnel should be able to book a time for a patient to visit a hospital instead of just recommending the patient to book a time themselves. This should relieve pressure from ambulance workers to solve less severe cases while providing better healthcare service to the patient.

When it comes to UI in the experience of the expert, healthcare professionals want explainable AI to be able to understand the decision process and make informed decisions themselves. In addition, an input system for AI should work alongside the general EMS Clinical Decision Support Systems (CDSS) such as NEWS.

Borgström, CSAM - Paratus demonstration
Highlights:

- Demonstration of the next to latest Paratus ambulance IT-system
- Flexible system, being able to adapt to any device size
- Live data collection from the whole healthcare chain, allowing for better data integration.

An interview was conducted with Jonas Borgström, product manager at CSAM, working on the Paratus ambulance IT system. CSAM has the advantage of having products in the whole healthcare system, allowing them to work with multiple kinds of data in real time and continuously utilise that in different ways.

A demo of one of the latest editions of Paratus was given (Figure.), showing how it currently gives live feedback based on the RETTS triage system. At all times the user can also access healthcare guidelines, based on what type of situation it is in that the user preselects.

During the majority of the interview time he showed us the demo of the system, which in general seems to have a high degree of usability. It was shown to be flexible, adapting to any window size provided, from desktop to tablet, to phone. The customer, for example, in different regions in Sweden can have different preferences for what contents should be a part of the system so the system is modifiable.

Bengt-Arne Sjöqvist, Professor of Practice Emiratus, industry veteran
Highlights:

- Several current systems, all working similarly, often with a touchpad-based interface.

- Sees potential in collecting data live from several sources, both structured and unstructured.
- Future of prehospital decision support should be able to be dynamic to facilitate as correct and early decisions as possible.
- Should be able to drill down and get more information.

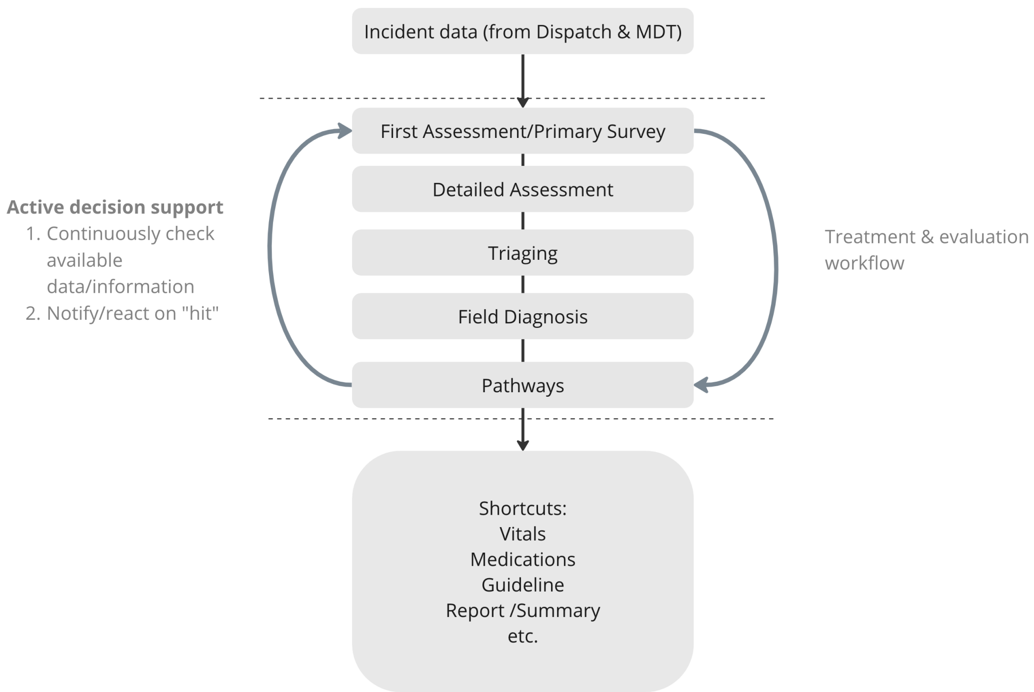


Figure 5.7: An example of Dynamic Risk Prediction implementation structure with added active decision support, based on an example from Bengt-Arne Sjöqvist.

Sjöqvist has extensive industry experience, a big part of the development of the Mobimed system, an IT system for ambulances that have become the industry standard and is widely used globally. When starting out, the goal was to do everything that was currently done on paper but digitally. Several other systems are currently on the market, but all of them share a similar interface, with main navigation on the side and with similar contents. The UI that is being used currently works at a satisfying level.

The current level of data access and the possibility to gather and evaluate data in real-time brings many possibilities to decision-making. Sjöqvist has always worked towards making decision-making in the ambulance quicker and more efficient. One major area of improvement he sees in current digital decision support systems is for them to be more active and dynamic. Current decision support follows a decision tree structure where one step leads to the next. A first initial assessment is made. Then, if no immediate emergency is found, the workflow moves on to a more detailed assessment, then triaging, etc. Instead, the future of healthcare should actively in real-time check the information, and notify the user when enough information has been gathered to make a decision, thus quicker and more correctly coming to a conclusion of what to do with the patient. Sjöqvist called this dynamic risk prediction, see Figure 5.7.

Time is often the difference between life and death or a much longer time for rehabilitation, thus more efficient and faster decision support systems would be a great benefit for patient safety. In addition, Sjöqvist sees the potential of such a system to allow the user to "drill down" on the provided information, to understand what causes the system to give a recommendation and provide sources to it. Thus it becomes more transparent and the user can gain a greater understanding of causes for a higher or lower risk prediction, especially if the system goes against the experience of the ambulance personnel.

During the interview, we were also given a quick demonstration of an older version of Mobimed to see how it was structured and how it handled information input.

Anna Bakidou, the PhD student in trauma algorithm
Highlights:

- Both transparent and "black box" algorithms were tested with similar success rates.
- Trauma has a high degree of triage error, making it particularly interesting for research.
- Model based on information that could be gathered from a prehospital setting.
- Not fully verified since it doesn't take distance to the hospital into account for under-/over-triage.

Anna Bakidou is a PhD student at the Care@Distance research group who has developed an AI model for trauma patients. There have previously been developed models for traffic-based trauma in the group, but her model is more general and attempts to work on all trauma cases.

A primary reason for Bakidou to be interested in trauma is that it can be very complex. The variety of cases is big, making it a unique problem to solve. There is also a big problem of both under-triage and over-triage, increasing the need for better decision support.

Various algorithms have been tested, from advanced neural networks to simpler and open algorithms. Both types performed similarly, making it possible that simpler algorithms are a better choice since neural networks are much more complex and thus lack the same level of transparency and explainability. However, since the algorithms haven't been optimised and variables could change in future development, more research is required before being able to be sure that the simpler algorithm works better in this context.

The models are trained and verified using the Swetrau database. It contains a variety of variables, such as what caused the injury and what body regions were impacted. The severity of the injury is also calculated using the new injury severity score. Much of this information is currently not collected by ambulance workers, but the model is based on what kind of information could be collected in the ambulance setting.

Currently, the model appears to be better on average than average at triage, compared to Swedish ambulance personnel. A major limitation of the model currently is that it doesn't take distance into account. Thus, if a hospital is too far away, the ambulance might decide not to transport the patient to a hospital, and this could make it appear in the system as an under-triage.

5.4 Requirements

Through the findings from our interviews and ethnography, we came up with a list of requirements for the system.

- **Integrate well into current and future systems:** Ambulance IT systems already exist on the market. The goal of Care@Distance is to bring their technologies to ambulances as soon as possible. Thus, creating an entirely new system is not feasible. To make the implementation of their system realistic, the design should be made in a way for it to be adaptable to existing systems.
- **Blend into the EMS workflow:** Obviously, for any system used in a professional environment, it needs to fit into the workflow of the personnel. The complex and ever-changing work environment of ambulance personnel makes this a big challenge. In addition, while AI is being developed and evaluated, it will need to work beside current triage systems.
- **Increase trust and understanding of AI:** AI is new and more complicated than existing triage systems, making it more difficult to trust and understand. The decisions made are often about life and death, making trust in the system is essential in order for it to function since it is still the personnel that is responsible for the decision.
- **Efficient and effective regarding stressful environments:** In order for the system to work, it needs to be used. This is currently not the case with many existing triaged based IT-systems, both according to our ethnography and a meta-study (Porter et al., 2020). In order for it to be used and not increase the cognitive load of the user, efficiency is essential. Thus, reduced steps to complete an action and easy overviews will always be paramount when making design decisions.
- **Versatile, support various kinds of situations:** Should be able to be used both in very acute and less acute cases. In addition, the system will be based on trauma. However, for it to be relevant, it needs to be a possible design for other areas as well.

For System requirements, We used the MoSCoW technique for system requirements to prioritise and distinguish what we must, should, could, and will not have in the design, see Figure 5.8.

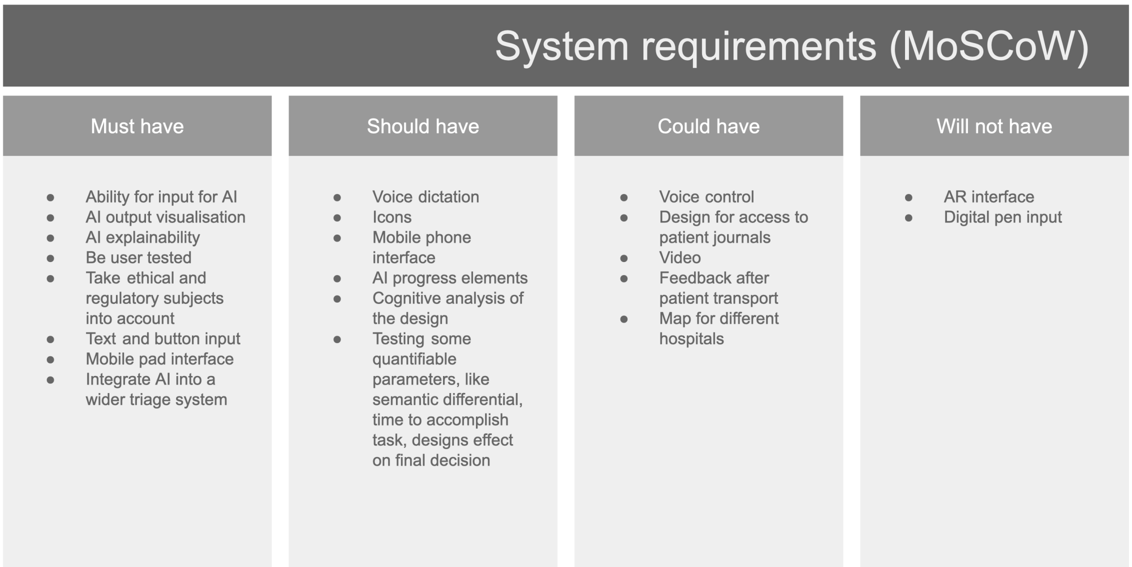


Figure 5.8: MoSCoW mapping

5.5 Prototyping

Throughout the project, feedback was gathered from stakeholders, users and experts, either through showing off our progress through online video conferences or through recorded videos. Thus we could gather information on what parts of the design were liked and which parts needed improvement.

5.5.1 Low fidelity

This section presents the result of the low-fidelity prototyping for this project divided into tangible prototypes and digital prototypes.

5.5.1.1 Sketch and tangible Prototype

We began with a tangible prototype where we structured, restructured, added, and discarded the design from the sketch, see Figure 5.9, as well as explored the actual size of the device by paper cutting, and experimented with different components that may fit within the UI concept. The tangible prototype provided hints of early stages of interaction that reflect the system flow and simulates the UI in the lower realism in a physical environment and allowed for rapid changes and testing of new ideas. Later on, the most relevant ideas were developed further into digital wireframes. See Figure 5.10.

5.5.1.2 Digital prototype

The digital prototype allowed us to explore the interaction of a digital device, but it also demanded more time to build. In this phase, we used the interactive prototype to communicate the ideas with stakeholders and gain their early feedback. The prototype was reformulated according to the feedback and the new findings.

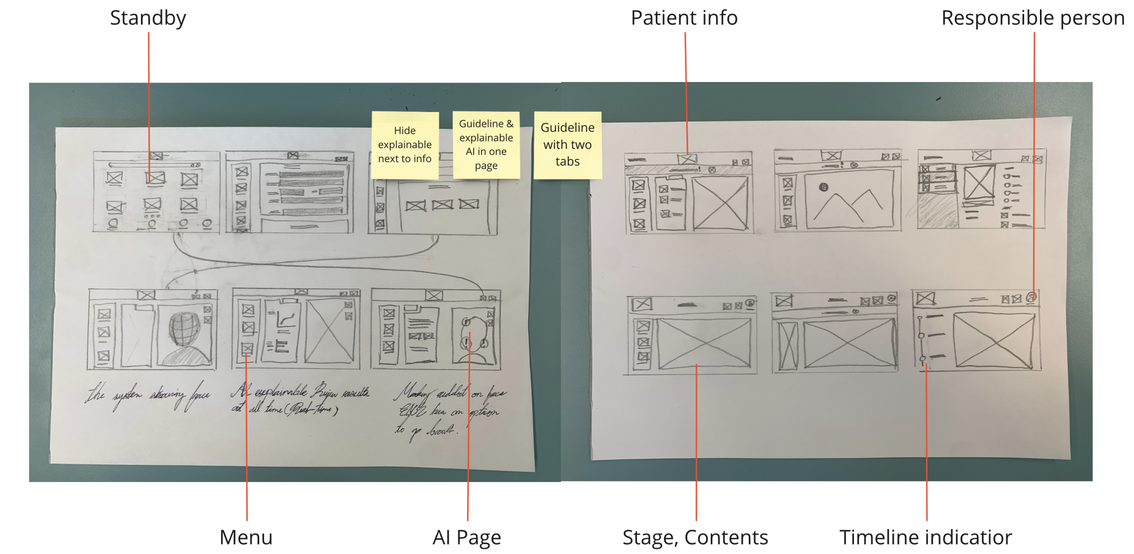


Figure 5.9: Sketching based on fictitious scenarios.

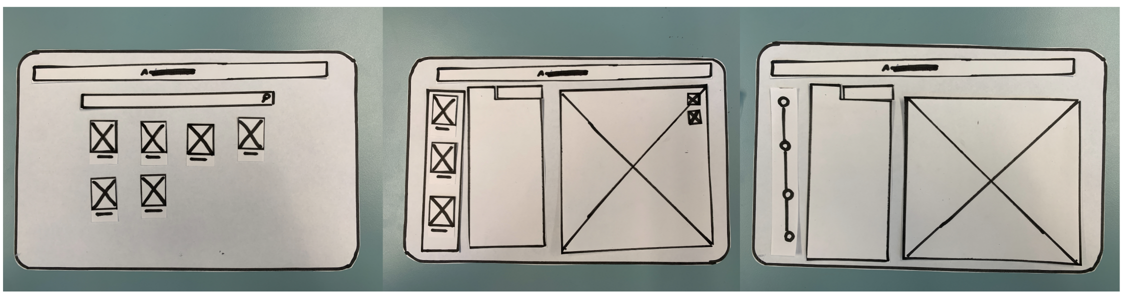


Figure 5.10: Tangible prototype, brainstorming and generating ideas.

5.5.1.2.1 Components They corresponded to four input categories: number entry, free text entry, and option picker. The categories were chosen based on what type of data EMS personnel currently need to enter for patient assessments. Each data entry method is based on a material design suggestion for an Android device; the guideline was developed by Google in 2013 (Material Design, 2023). However, We have adopted some elements to this project for ambulance services. For instance, The button and the free text entry require a bigger size than the standard mentioned in the Material Design due to the shaking of a moving ambulance; dark colour needs stronger contrast to notice in stressful environments.

Number entry Figure 5.11

- Keypad, Allows one to select a subject and enter values through a keypad.
- Dictate, Allows the users to add input numbers during hands-full tasks.

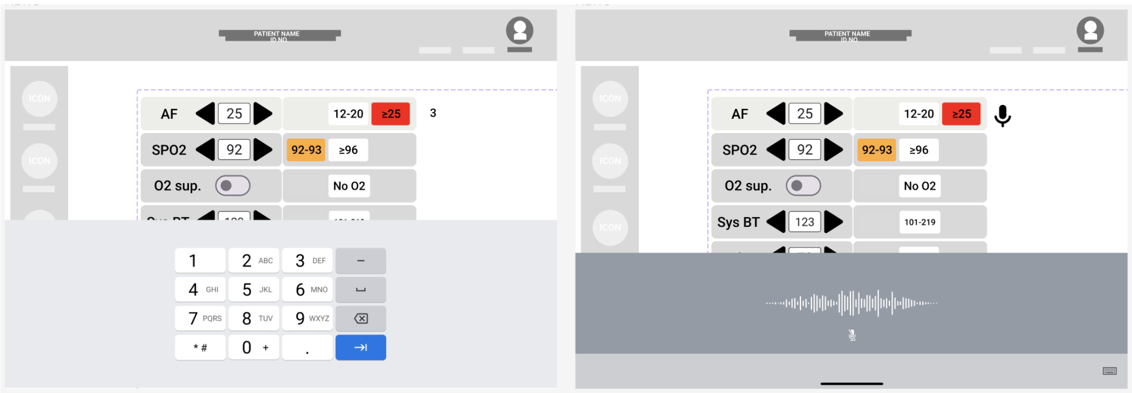


Figure 5.11: Example of number entry solutions.

Free text entry Figure 5.12.

- Keyboard, To allow users to select the field in which they want to enter data and use the digital keyboard to do it and a physical extension keyboard is available if required.
- Dictation, To provide an alternative option while physical contact with the device is not possible. Users can dictate by simply pressing the microphone icon or using word commands to activate it.

Option picker Figure 5.13

- Checkboxes, Typically similar to a conventional checklist on paper, allowing the user to check on each box by tapping the icon or the text connected to it.
- Dropdown menu: A solution to hide the options under a menu and reduce required space. It allows the user to click the dropdown icon and select it from the menu.
- Chip: Allows the user to make a selection, filter content and trigger action. The chip can show multiple interactive elements together in the same area.

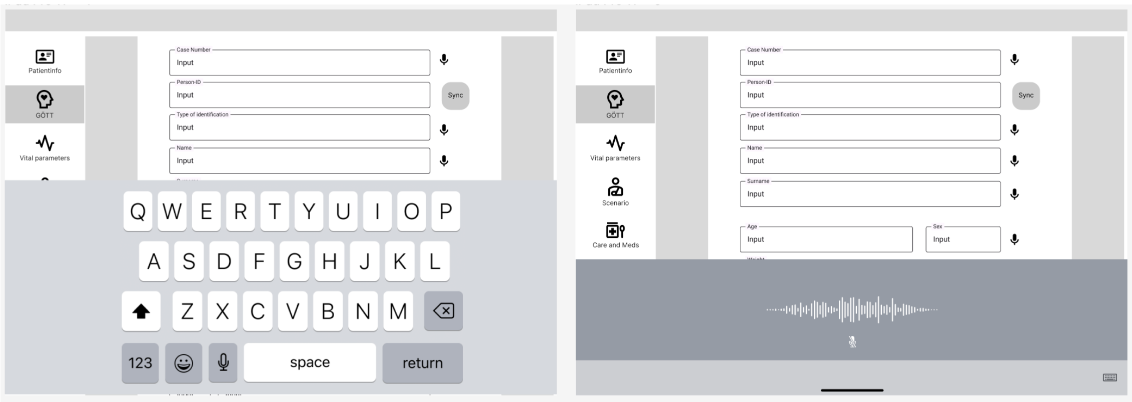


Figure 5.12: Example of free text entry solutions.

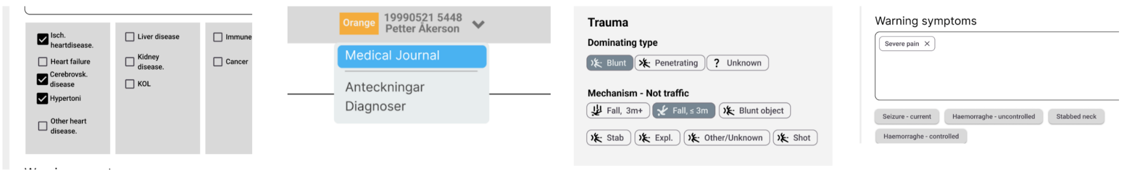


Figure 5.13: Example of option picker solutions. From left: Checkboxes, Dropdown menu, Two versions of input chips

5.5.1.2.2 User experience (Emotions) Another brainstorming session was set to discuss the UX aspect of the design. For example, to understand and have an overview of what emotions to evoke or what dynamic should be emphasised to motivate positive experiences for the users. Therefore, the Mechanics-Dynamics-Aesthetics (MDA) framework was adopted in this project to gain an overview of the relationship between interaction and emotions and to know the needed mechanics for enabling it. See Figure 5.14.

5.5.1.2.3 Physical format The design considers adaptive design as the main, user interface adapted to different screen sizes. The EMS personnel use multiple devices according to the circumstances and situations. However, this project is mobile tablet-oriented based on the result from the exploration phase that tablets are used the most in the ambulance.

5.5.2 Wireframe

The wireframe was presented in a neutral tone with a grey scale but with good enough contrast and well enough for communication. However, the focus was on how to display the AI in a comprehensive way in the ambulance setting. A few suggestions were granted and allowed to combine ideas (e.g., cards vs chart) Here seen as both an individual page in Figure 5.15 or as an overlay in Figure 5.16.

Besides the main focus of AI, for input to the system we have digitised the pen & paper acute journal utilising WEST, which is being used in the VGR today and is used by the EMS personnel in making prioritisation decisions for the patient cases,

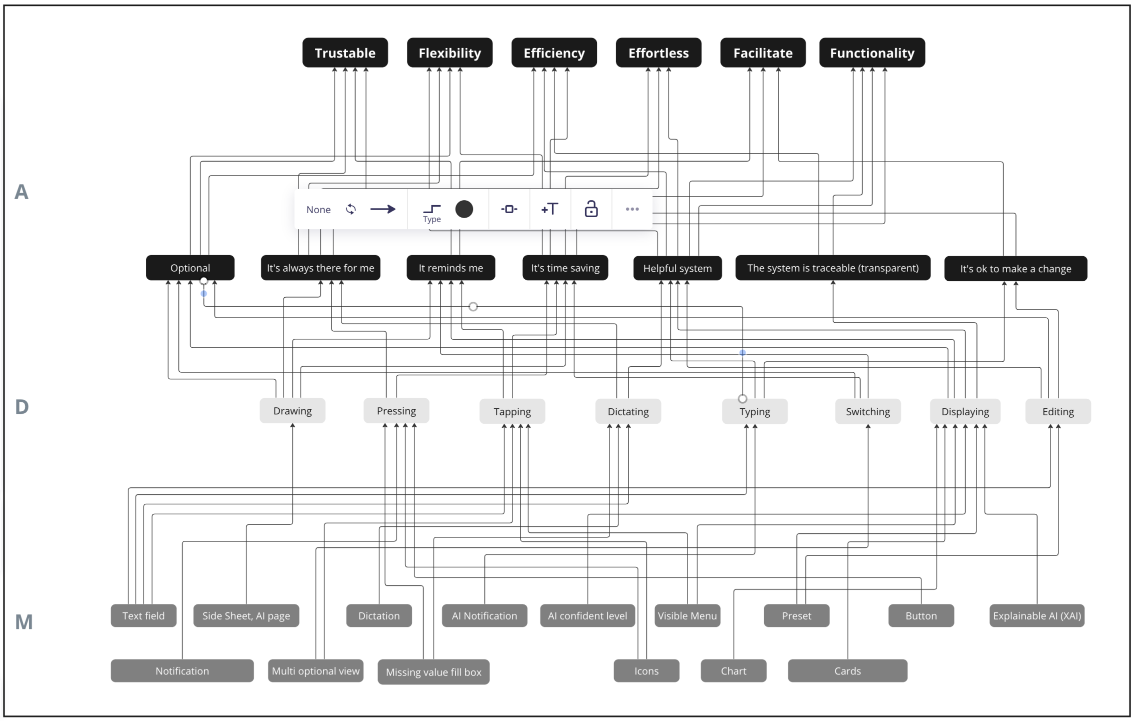


Figure 5.14: Diagram of MDA framework

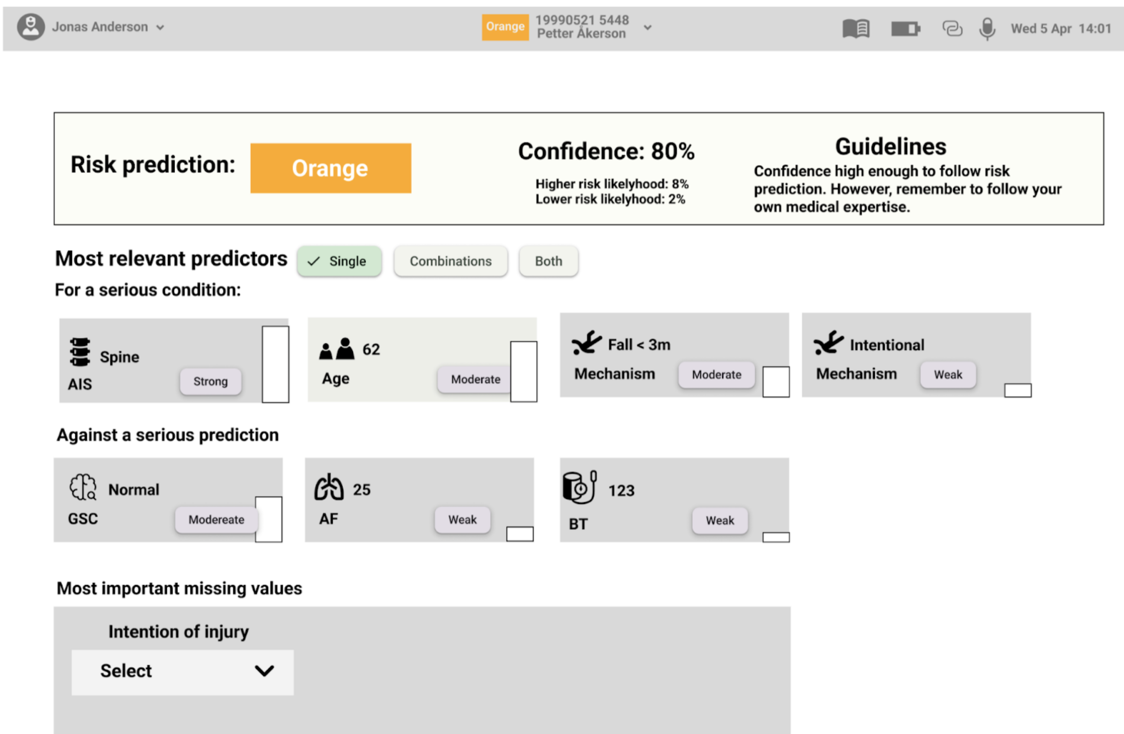


Figure 5.15: Early wireframe AI presentation with multiple cards representing different values.

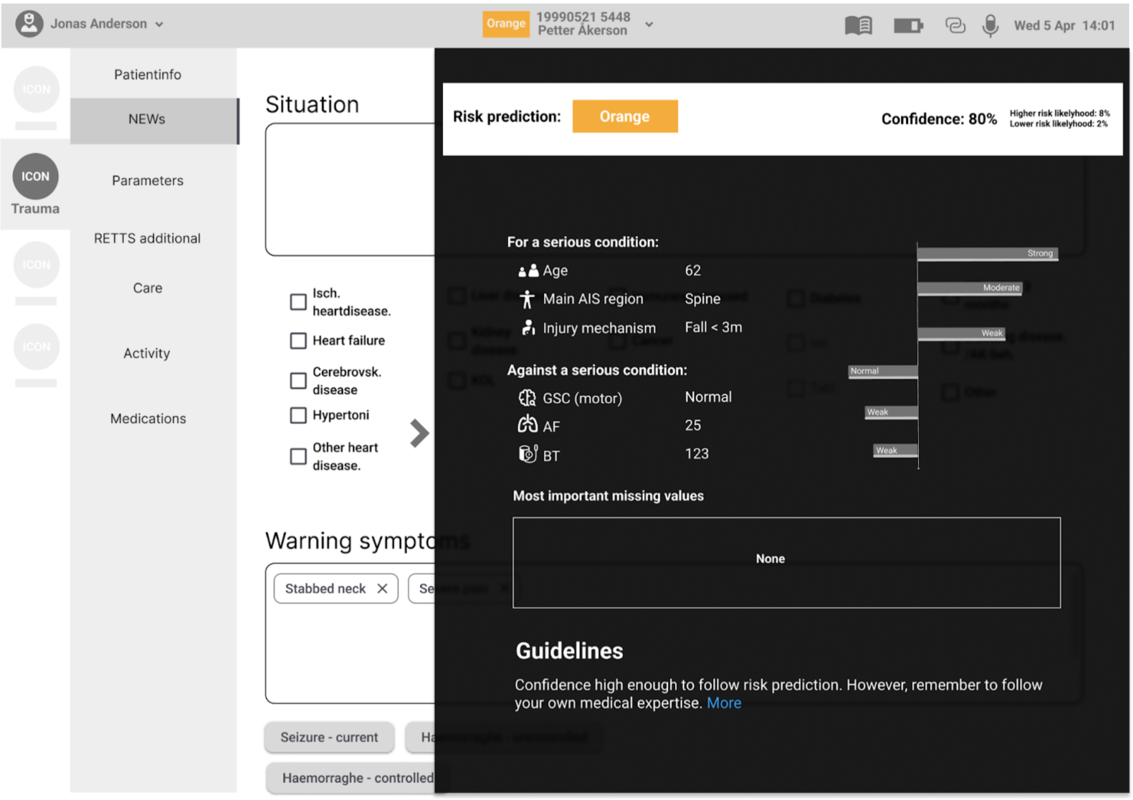


Figure 5.16: Early wireframe AI presentation overlay with a chart comparing different values.

see Figure 5.17. We reformulated the design a few times. This allows for a more complete workflow for the nurse, with both input into the system, and then the AI results as output.

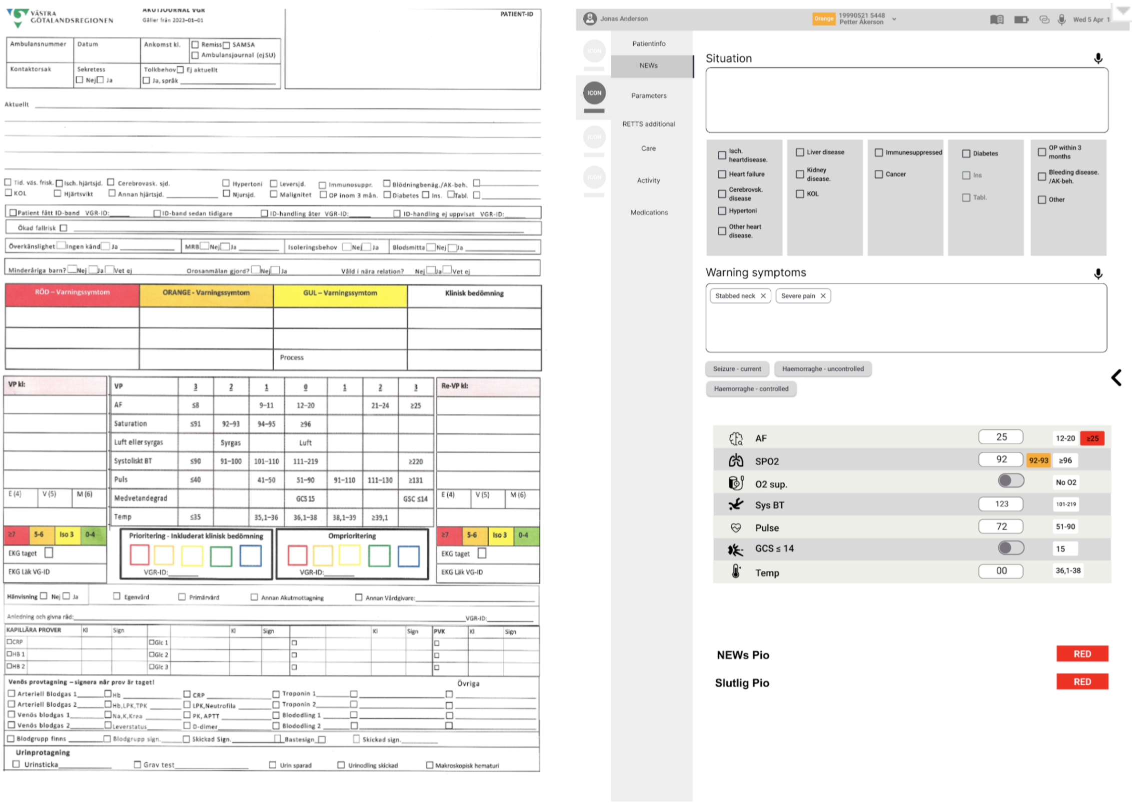


Figure 5.17: The WEST paper journal and an early digitised version of it. For the full-sized paper journal, see Appendix A

The wireframe strategy testing not only allows the project to move forwards faster but also enables us to make several better adjustments in the design by trying this early prototype on the actual device. Unfortunately, things were not as smooth as expected; at this stage, we did not have access to the actual device yet, but we speculated using Figma software in the testing mode and continued development while aiming attention to reliability and functionality to make the system more efficient. This solution aimed to solve the following; 1. to solve the issue with misinterpreted handwriting in the current paper journal, 2. to include most information in one device, 3. to reduce time consumption by fewer interactions or a number of clicks, 4. to increase visibility and reduce overwhelming information, 5. to emphasise the most critical information.

5.5.2.1 Decision support presentation

The Decision support consists of four main parts:

1. The top bar displays the overall risk prediction using the same colour codes currently in use by ambulance services. It also includes the confidence level

of the decision support, which can increase or decrease depending on the information input.

2. The guideline is a part where the decision support provides feedback and recommendations. It also provides further details accessed by pressing "more" as an option.

3. Decision explainable AI (XAI), this part is a passive presentation of what the AI decision is based on. The view can switch into two modes, both displaying the same info. The card mode shows all predictors that the AI was using for the risk prediction. It shows them in two rows, one for a serious condition and one against a serious condition. They are sorted from highest to lowest contribution. The chart view is more limited in that it only shows the top three predictors of for or against a serious condition but allows for a quick overview of their individual contribution. A horizontal chart was chosen instead of a vertical one to be able to contain all the required information about the predictor on the side.

   An exploration about the possibility to drill down information on the predictors was explored, see Figure 5.18. This could be important for learning the logic behind the system. However, after the ethnography research where interaction with the system was very quick and after asking a nurse for feedback, this feature would seem to be unlikely to be used during triage. This, combined with the added knowledge acquisition this feature would require and time constraints, the feature was not investigated further. Instead, the functionality was imagined to be moved to the summary page of the journaling system where nurses would have time to investigate.

4. Missing value allows the user to input information that the AI thinks is the most important for increasing its confidence level in the prediction. It allows for quick input without going through the whole system or changing views.

Figure 5.19 displays a result wireframe as a whole. It includes pages that develop according to the imagined scenarios and builds for usability testing. There are several thoughts from this process on how the system should look.

1. The AI page should be able to assist at any time.

2. The AI prediction should be visible at all times.

3. The top bar should present the patient's brief info and the EMS personnel responsible.

4. Pages that are long must be able to pan down.

5. More visuals (Icons) should be added to reduce the focus on reading text.

6. Having much space in between content helps against overloading users.

7. Size of text and icons needs to be bigger for the ambulance environment.

   The more developed version of the wireframe, see Figure 5.20, was used to communicate with the stakeholders and an experienced EMS nurse. A video

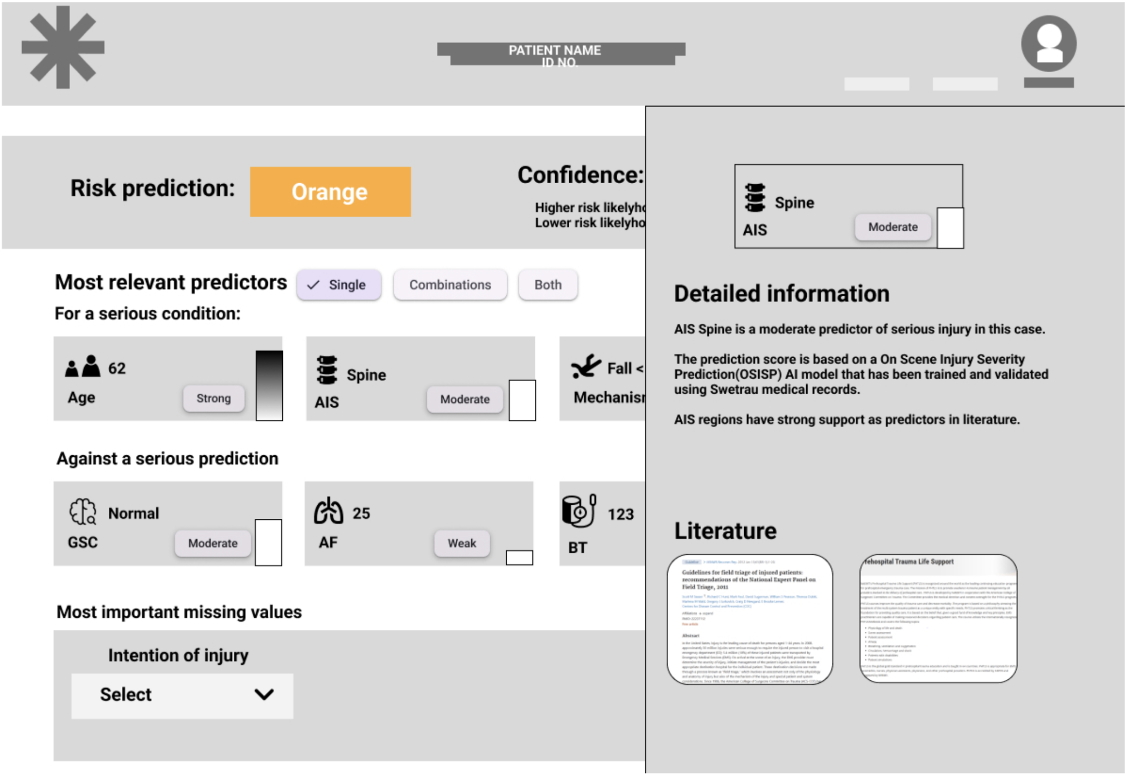


Figure 5.18: Drill down functionality during triage.

demonstration for a workflow was recorded using voice and screen recording and was then sent to gather feedback. The same video was used in supervision as well.

Points from the feedback included the need to create a versatile design and have multi-optional choices due to the requirement of individual nurses in the ambulance context. The feedback included discussions about UI elements like how to present the confidence rate for AI risk prediction (e.g., score system, scaling system and dot system).

Another point was about the XAI presentation, where we discussed how to visualise and present it. Two options were presented; the cards view or the chart view. The card view gathered more positive feedback, but it was concluded that both options should remain for the ambulance personnel to decide.

5.5.2.2 Visual design

Visual design covers several important aspects which increase usability and positive user experience by establishing a visual hierarchy: the most probable should be most prominent, position (top, middle, and bottom), visual attribute (colour, size, space, typography, readability, and aesthetics), alignment (use of the grid system to minimise visual excise and group object) and balance (left and right of UI).

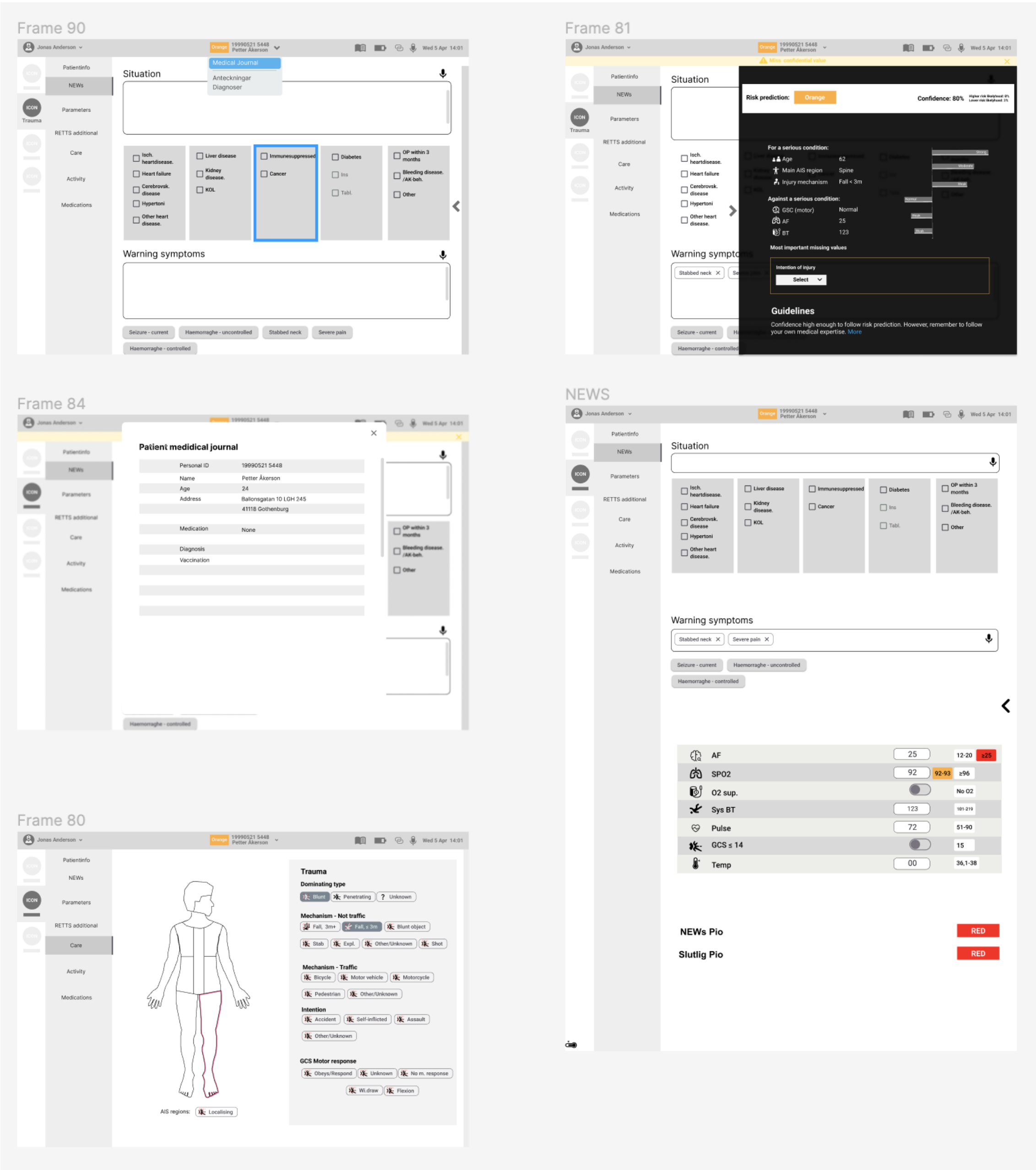


Figure 5.19: The wireframe in different stages in the system.

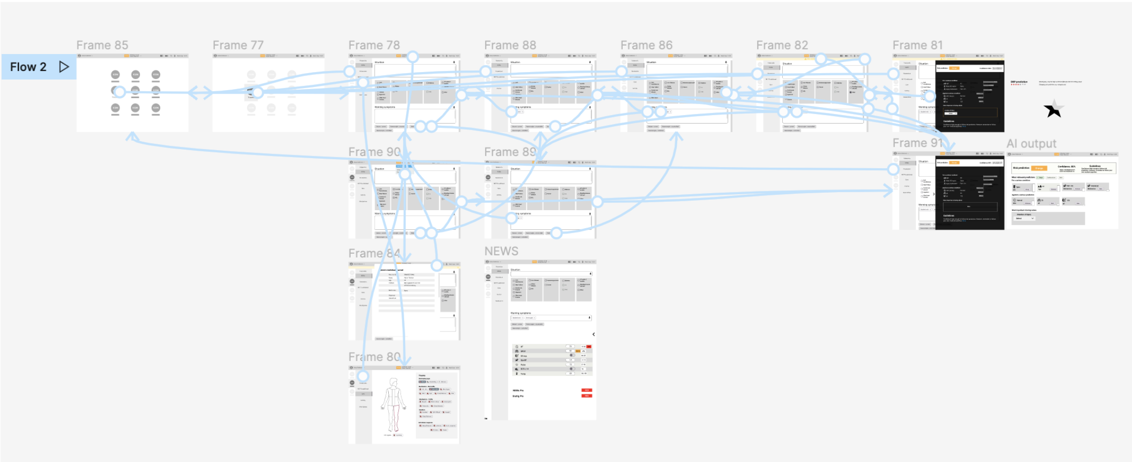


Figure 5.20: Interactive prototype flow created in Figma

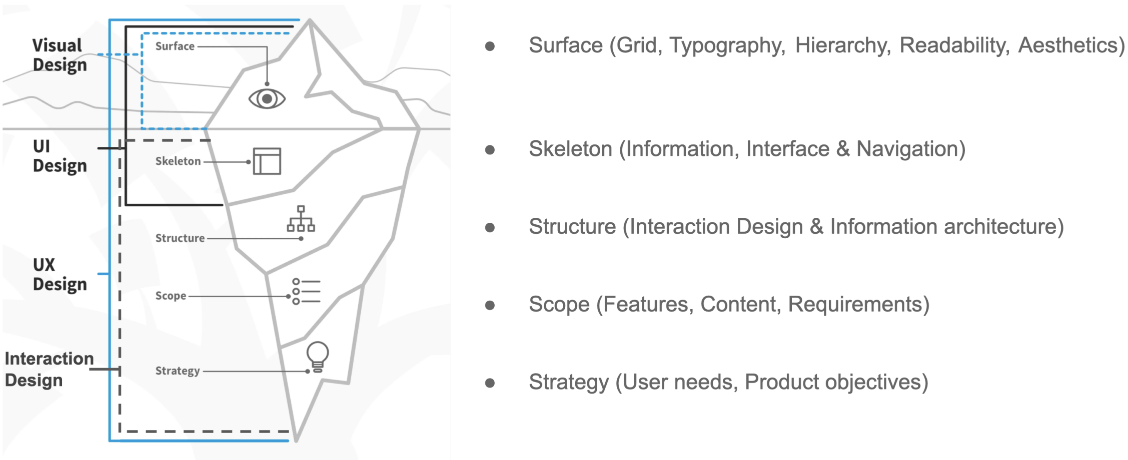


Figure 5.21: The graphic describes UX design, including Visual Design. Adapted from Interaction Design Foundation (2002).

5.5.2.2.1 Accessibility In the following design approach, we focused on inclusive design (Imrie & Hall, 2003). aim to increase the design quality in respondent behaviour to the needs of individuals.

The design intends to consider information visualisation, which considers some adjustments toward ambulance services based on information visualisation perception for design (Ware, 2019). Thus, the mood board was created to brainstorm and generate ideas for visual communication inclusivity; colour, lightness, brightness, contrast, and constancy. See Figure 5.22

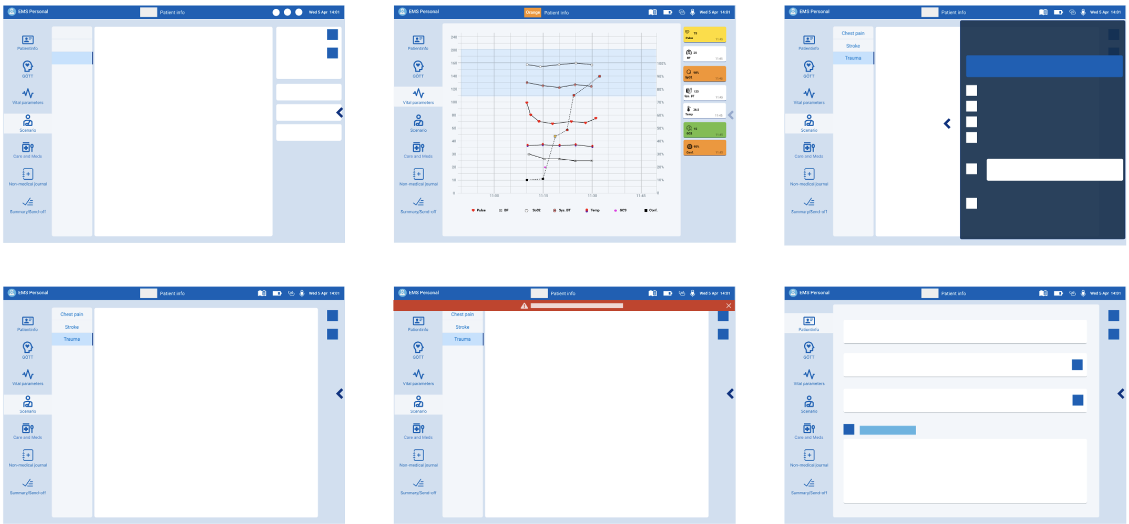


Figure 5.22: A mood board presenting early visual interface and potential design.

5.5.2.2.2 Interaction This part is about improving the interaction to ensure the design has a high functionality standard for an ambulance and enhances the user's positive experience. First, however, avoiding switching the page for the leading AI page is crucial. We are considering the data manipulation loop into account. According to Ware (2019), At the lowest level is the data manipulation loop through which objects are selected and moved using the basic skills of eye-hand coordination. Delays of even a fraction of a second in this interaction cycle can seriously disrupt the performance of higher-level tasks.

The option in the AI page to gain time reduction and voice switching page to reduce the reload screen, e.g. where the card is located, the user can continue to pan the card to beagle to view the rest of the card. See Figure 5.23. Another general UI included Icons, checkboxes, a change in surface colour, or a combination of highlighting to minimise mistakes and have better control over the interface.

5.5.2.2.3 Layout The layout, see Figure 5.24, is designed to fit the different screens. Therefore, it should be automatically adaptable for any device. The following figure demonstrates an example of a tablet size typically used in an ambulance in Sweden. Based on the data from the exploration phase, tablets are the most common hardware to use for IT systems; therefore, the design is tablet-oriented. Furthermore,

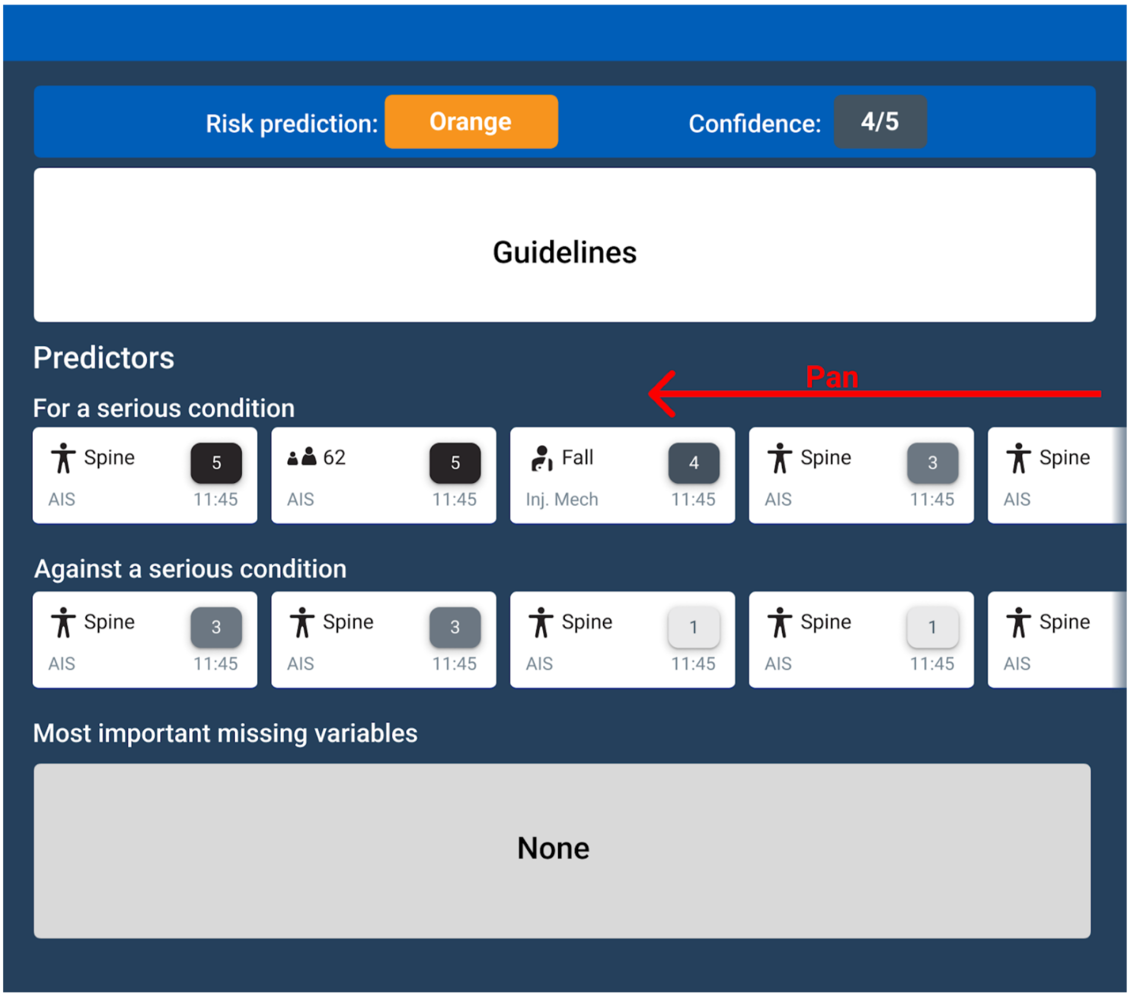


Figure 5.23: AI Overlay. More cards can be seen by panning from the side

the layout was created to consider how the spacing and the parts of the layout could work together in harmony with composition positioning.

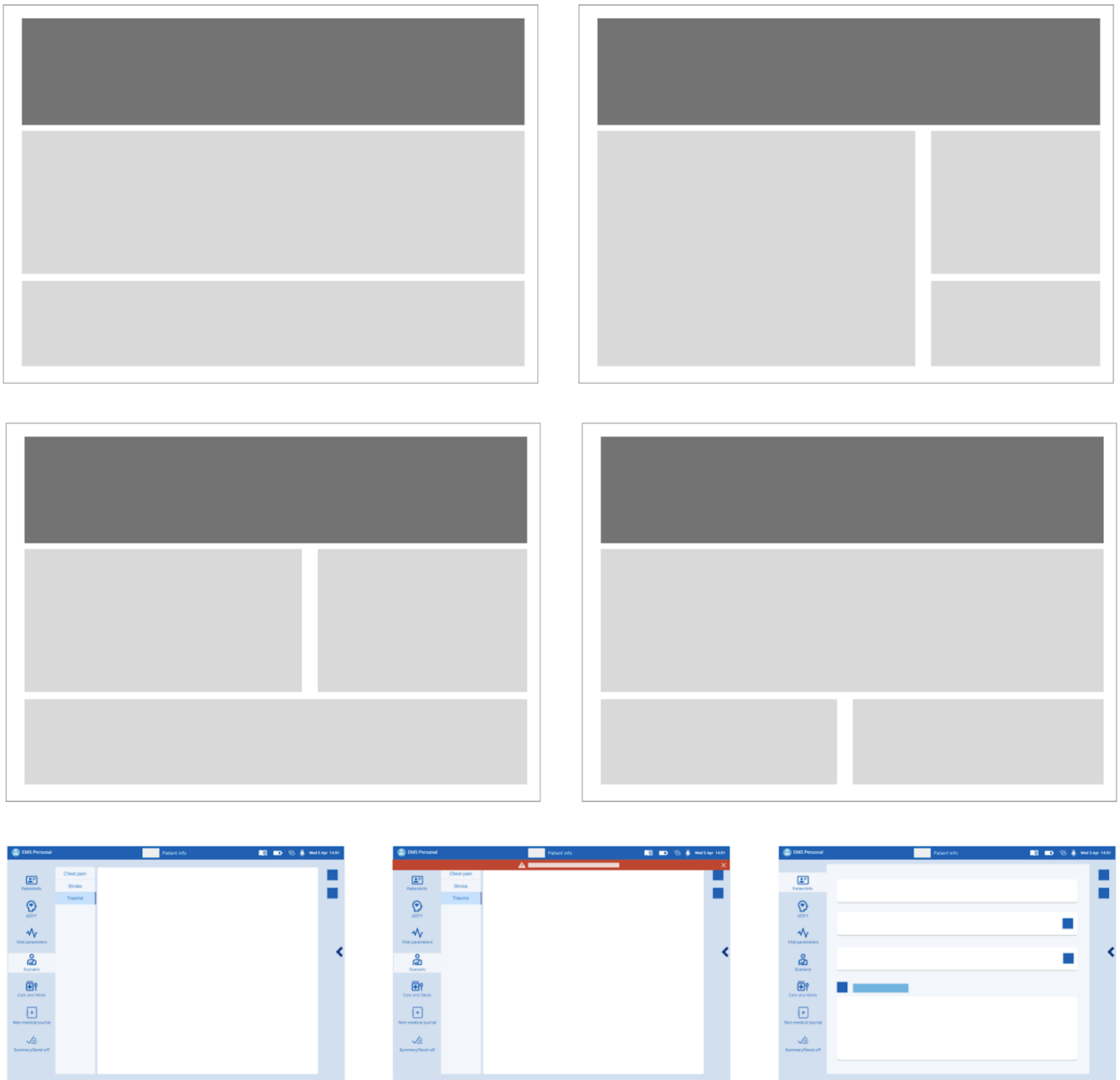


Figure 5.24: Sketching ideas for the design layout

5.5.2.2.4 Styles

- **Colour:** The selection, see Figure 5.25, follows a guideline from The National Health Service (NHS) (2023). The NHS is the publicly funded healthcare system in the United Kingdom. NHS provides high confidence in healthcare design tested on the end user. In addition, the selection of colours helps people recognise and trust the services (NHS, 2023). For this project, some colour was picked from the options provided via the recommendation of the NHS, which has been tested already, but also by our choice where it has to adapt to the ambulance context. We use colours to help users prioritise and distinguish or recognise differentiated information. Text and background colours are designed to meet accessibility needs according to the NHS (2023) guide, considering contrast, space and colour choices will increase accessibility for people with colour blindness; text describing the colour was also added for clarification where it was deemed most important to avoid confusion.

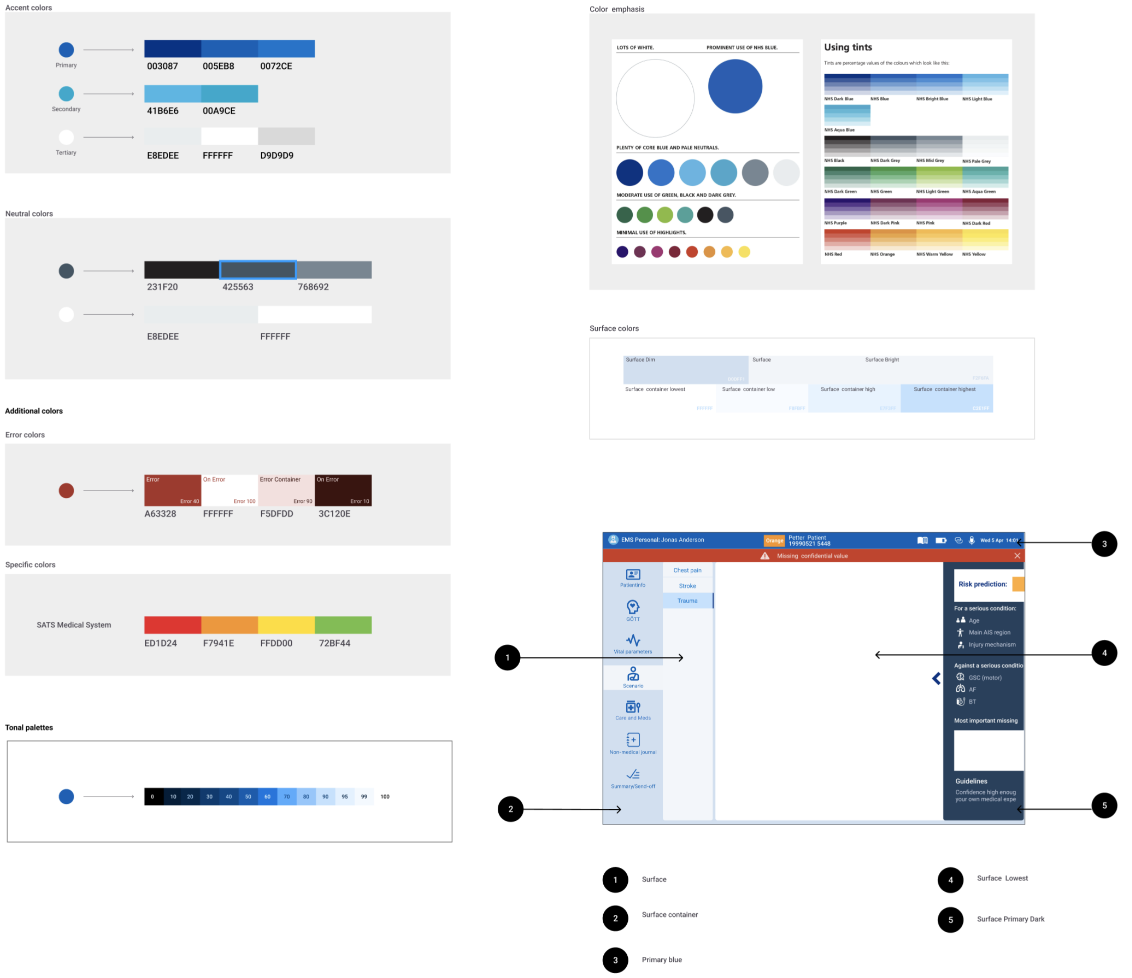


Figure 5.25: The selected colour pallets for ambulance work service.

- **Typography:** The font of choice in this project is Roboto Flex, see Figure 5.26, having a fluid range of weights, from extremely thin to highly bold styles in larger optical sizes. The font is presented in five type styles defined by five roles: display, headline, title, body, and label. These new roles are more descriptive, allowing straightforward applications across a broader range of use cases (Material Design, 2023). The Roboto font is the standard for common Android devices. Since the project is meant to be for versatile use in the ambulance environment work, Roboto tends to be the right font for this.

5.5.3 High fidelity

Information, interface and navigation design

The overall layout of the design can be seen in Figure 5.27. It follows the pattern of metaphors, idioms and affordances from the current trends in the market, SOTA.

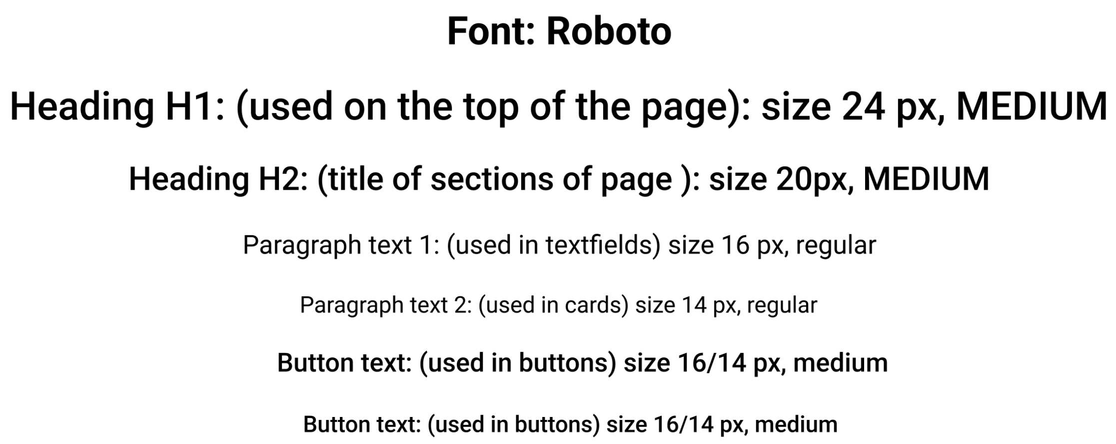


Figure 5.26: Font Roboto, used throughout design

The navigation bar to the left is a prime example of this, being used in both the Paratus and Mobimed IT systems. During our visit to the ambulance in Halland using an older version of Paratus which only uses text for the navigation bar, we noticed that even experienced users had to read through the menu to find the right button. Icons combined with text can make interaction significantly faster (Majrashi, 2022) while only using icons can be risky depending on how strong the association of the icon is with the contents of the page (Islam, 2015). The only accessible pages of the prototype are WEST, Vital Parameters and Scenario. The rest have already been designed repeatedly by other companies and don't add as much to the diagnosing process, so they have been left out. Summary/Send-off could have benefited from design related to the AI decision support, where more detailed information about the AI risk prediction could be displayed that there would usually not be time for during triage. However, this was not the focus of this project and thus was outside of the scope.

There is a top bar at the top of the page. This is somewhat similar to the Paratus system, though in this case, information about the patient is moved to the middle. To the left is information about which user is logged into the system. The middle starts with the current prediction of the AI algorithm, followed by the id-number and name of the patient. The AI indicator and id-info are close together, utilising the Gestalt law of proximity, showing that they are connected (Todorovic, 2008). Clicking on the patient allows for access to an overview of their journal and information from the SOS call. Finally, to the right, there are icons showing guidelines, battery life, two rings showing if the system is connected and voice control followed by date and time.

Currently, the WEST page is shown. It is using all the parts relevant to diagnosing the patient from the current paper journal used in the VGR ambulance. This knowledge has never been digitised before. The page uses the same general order of the paper journal, though with the NEWS vitals to the right of Warning symptoms instead of below it. This provides a complete overview of WEST without any scrolling or interaction required.

It starts with a simple text field, allowing the user to input information about the circumstance of the situation. A similar text field can be seen below. To the right, there is a microphone icon for dictation. Due to the limitations of Figma, the text input in this prototype is automatic. Otherwise, it would use either the pads screen keyboard combined with autofill, or an external keyboard.

The situation is followed by checkboxes with preconditions, categorised and ordered into different boxes and utilised using the Gestalt law of common region (Todorovic, 2008). A tiny outline is used to help guide where to place the finger. In the paper, they are ordered as two lines, left to right, with no clear categorisation. When trying out that layout, it made certain preconditions appear more important than others and made it appear harder to read. Thus, this vertical layout was kept. While not as effective space-wise, it should make the page easier to read. Dropdown menus could have been used, but we wanted to keep them all available at the same time, to make interaction quicker and easy to correct errors, while also being there as a reminder for the user as to what preconditions might be important for the case.

Warning symptoms in the paper journal need to be categorised by the users themselves. Here, the symptoms are already categorised into different colours, so the user needs to only write or select them to get information about how serious they are. The suggestions shown are based upon the previous information that has been put in, such as SOS call, age, situation, other warning symptoms, etc. As more warning symptoms are put in, it updates with more relevant suggestions. This is to both remind the user what symptoms to look for and also allow for quicker interaction than writing or dictating.

Finally, in the bottom right, it displays vital parameters based on the NEWS section of the paper journal. For each line, an icon is followed by a text description, digits indicating the current number, and a symbol presenting if the information has been retrieved digitally from the device recording it. On the right, the middle number indicates the non-significant parameter numbers. If the current number is outside of the parameter, it is shown to the left or right depending on if it is lower or higher. The colour indicates how far outside the parameter it is, with yellow here indicating step one outside non-significant parameters. Finally, NEWS PRIO shows the final NEWS indicator, here Green since there is not a single parameter far exceeding normal values and not enough parameters exceeding normal values by smaller degrees.

Scenario
When clicking on Scenario, the user is taken to the page shown in Figure 5.28. For demonstration purposes, some information has already been selected. In the imagined scenario, the user has just selected “Motor vehicle” as the injury mechanism, and gotten a notification that the AI is ready to give a prediction. The notification is animated in that it pulls down from the top bar, to guide attention. The top bar AI indicator has also changed, now showing Yellow to indicate the changed prediction.

The scenario page is imagined to be able to accommodate several different kinds of scenarios. A submenu to the navigation bar has been made, which allows the user to select several different scenarios with specific inputs for that scenario. This is inspired by Mobimed which has a clinical pathways page, allowing for specific recording of

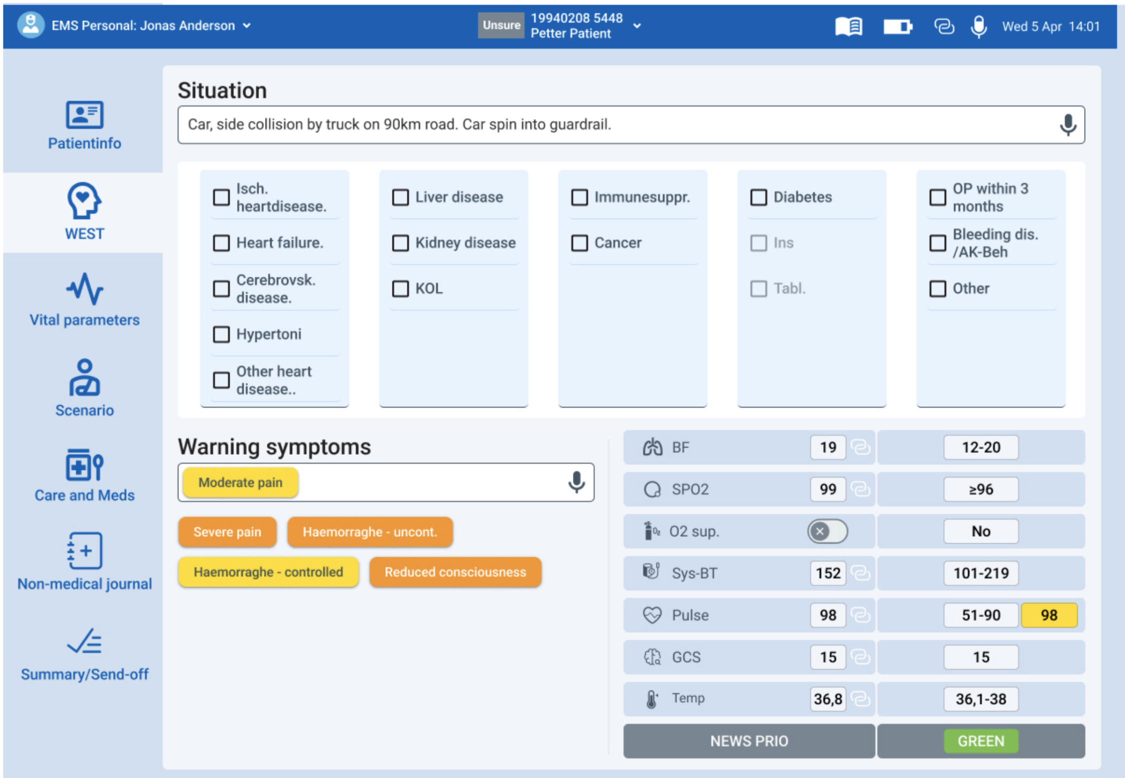


Figure 5.27: WEST page of the high fidelity prototype, based on the WEST paper journal for triage decision support

symptoms and treatment guidelines. The prototype has trauma preselected, with the system basing its selection on information received from the SOS call. This makes it so that the user doesn't need to search the menu and click it, but still allows them to select a different scenario if the system is wrong.

All inputs on this page are based on the model developed by Bakidou (2023). Different AI models can require different specific inputs, and the Scenario page allows for navigation of several algorithms, here with stroke and chest pain as examples, with a model for the former currently being developed at the Care@Distance research group and the latter used by Andersson & Maxstad (2022).

In the middle, you have Abbreviated Injury Scale (AIS) regions as a part of the AIS system to measure the severity of trauma injuries (Greenspan, McLellan, & Greig, 1985). The body is created using standards, with the main body from the VGR paper journal (See appendix) and the head from Bieler, et.al.(2021). Most regions are self-explanatory. Both the left and right sides count as the same region. The head is divided into both face and head injuries. Clicking on a body region selects it. Originally, an indicator was added below the body, showing what regions are selected. Clicking on it also allowed the region to be unselected. However, due to how Figma works, adding this interaction to the prototype was not feasible (Figure 5.28).

The right side contains the rest of the inputs for Bakidou's model. Only one button in each section can be selected at the same time. The buttons are input chips based on material design(Google, n.d.) but with increased sizes to accommodate the ambulance work environment. All inputs in the different categories are ordered based on how common they are, with the accident being the most common intention as an example, based on Bakidou, et. al., (2023). The reasoning is that this should reduce the time to read through the options most of the time, increasing efficiency and reducing cognitive load.

The first category is the motoric response part of the Glasgow Coma Scale(GCS) (Teasdale, et.al., 2014) utilised to measure consciousness level. This order is based on ambulance guidelines, where GCS is the most important variable for a serious medical situation out of this list (Region Stockholm, 2021).

Injury mechanisms have two categories, traffic and other. They can be switched between using tabs from material design. Here, Traffic is preselected based on the SOS call to avoid unnecessary clicks. While SOS call is quite limited in providing the information, it should be able to give enough information to differentiate the injury mechanism between Traffic and Other.

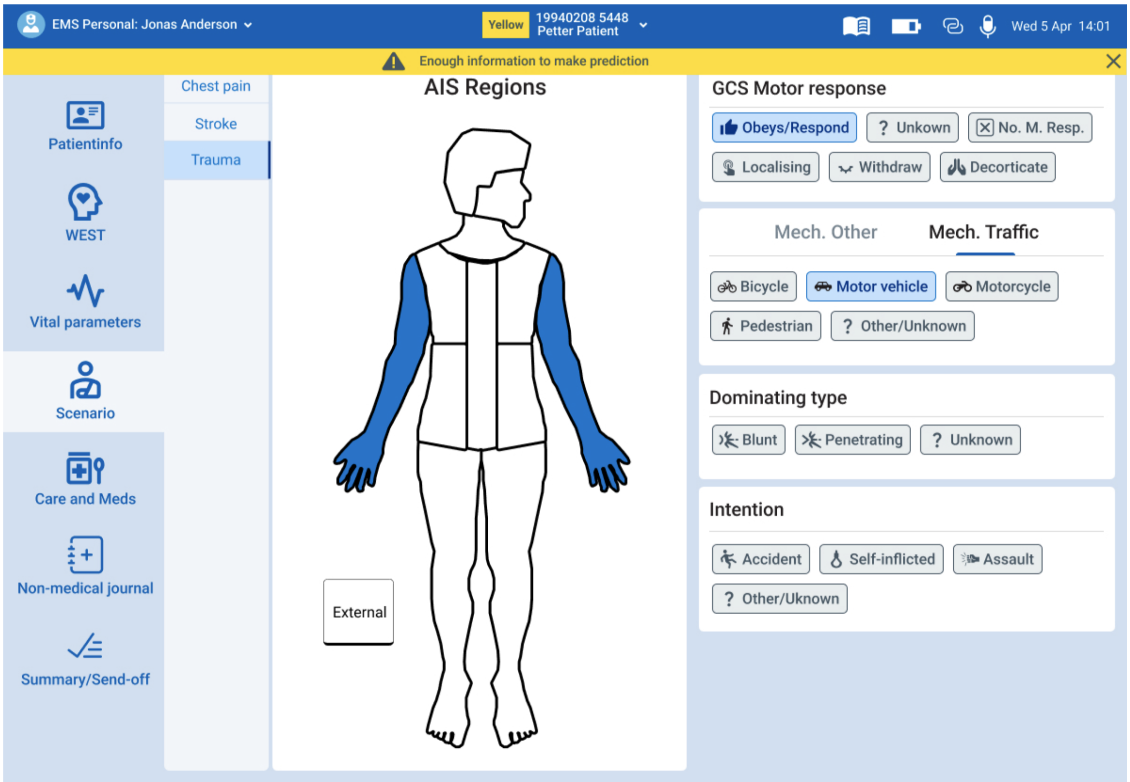


Figure 5.28: Regions as a part of the AIS system to measure the severity of trauma injuries.

AI Output
If clicking on the notification, the AI indicator on the top bar, or sliding from the right anywhere on the page, the AI output slides out from the right, shown in Figure 5.29. Sliding back on it or clicking outside it is closed again. This way, AI output

is always available, no matter which page you are on, and you don't need to stop what you are doing to get to the page. The page is divided into three main sections, prediction and guidelines, predictors, followed most important missing variables.

The first section provides the most important information, and for some cases might be all the user needs to see. It contains the risk prediction, confidence level and guidelines. The guidelines here are just that the confidence is high enough for the nurse to follow them and to use their own expertise. They could also contain different guidelines, related to the case, such as recommended treatments related to the case.

Predictors show the predictors that contribute to the confidence level of the algorithm, against and for a serious condition. These are card based, inspired by Paratus. They are ordered from left to right, starting with the most important contributor to the prediction. Sometimes, there can be more predictors available than there is space for, and then the individual lines are scrollable from the side. Confidence contribution is displayed with a number from one to five and uses neutral contrast levels to make it easier to distinguish the contribution level quicker. Colours are not used to avoid biassing the interpretation of the predictors. Contrast is not the most precise method of data visualisation (Cleveland, & McGill, 1984), but considering the data is ordinal and combined with the numbers, it was deemed good enough.

The predictor cards should be intractable, allowing one to change them and giving very quick information as to why they are important. This was investigated briefly in the low-fidelity prototyping phase but due to time constraints was not developed further for the testable prototype (Figure 5.29).

Clicking in the top right corner of predictors brings up a second way to view the top three predictors for and against, shown in Figure 5.30. This is an alternative, slightly more precise view of how much they contribute to the confidence level. The choice of which to show is left up to the preference of the user. The exact measurement of contribution is not really that important, but knowing which variables are most important and which ones can likely be ignored should inform the user of how the AI came to its conclusion.

Finally, the most important variables allow the user to add additional variables that the AI currently lacks but deems important, see Figure 5.31. A dialogue window appears, creating a similar input layout from the corresponding page that is taken from. Clicking outside it or making a selection removes the dialogue unless there are multiple possible selections, in which case a Submit button is added, as in the case of AIS regions. Selecting an input quickly makes the input be highlighted to confirm the selection before removing the dialogue.

Vital parameters, shown in Figure 5.32, are heavily inspired by the same page used in the Paratus system, with some changes. A highlight is added between 70%-100% to make it easier to see the most important percentage areas. All icons are symbols that relate to the parameter they are corresponding to instead of being arbitrary. Confidence level has been added as well, also displaying the colour-rated seriousness prediction. Having two different Y values, absolute numbers and percentages, might appear confusing to casual users. According to CSAM/Paratus's Jonas Borgström,

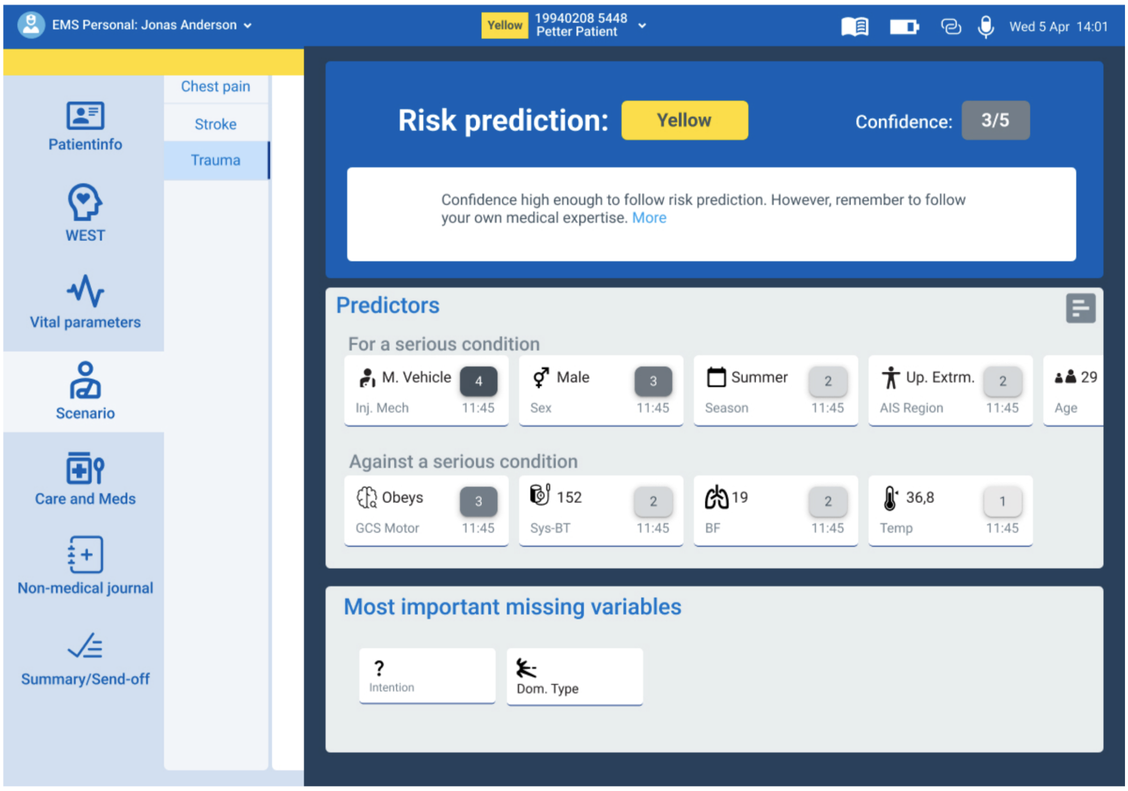


Figure 5.29: AI output overlay. The predictor cards should be intractable, allowing one to change them and giving quick information.

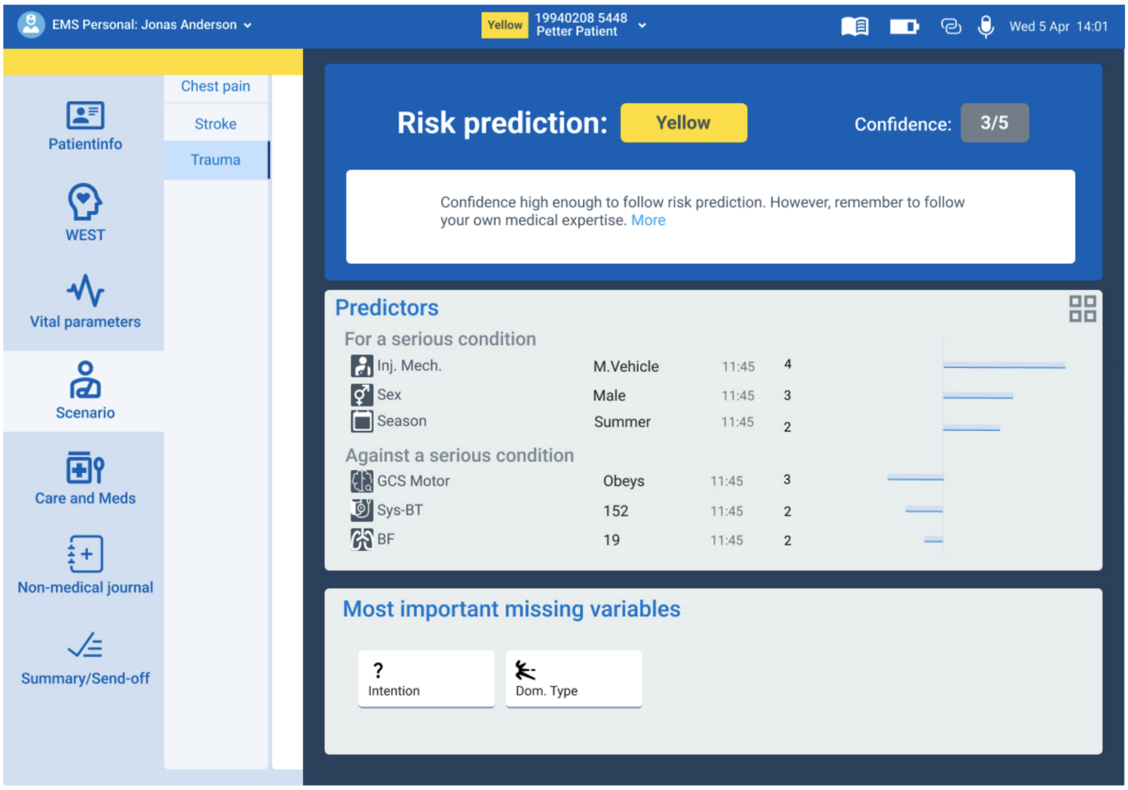


Figure 5.30: A second display to view the top three predictors for and against.

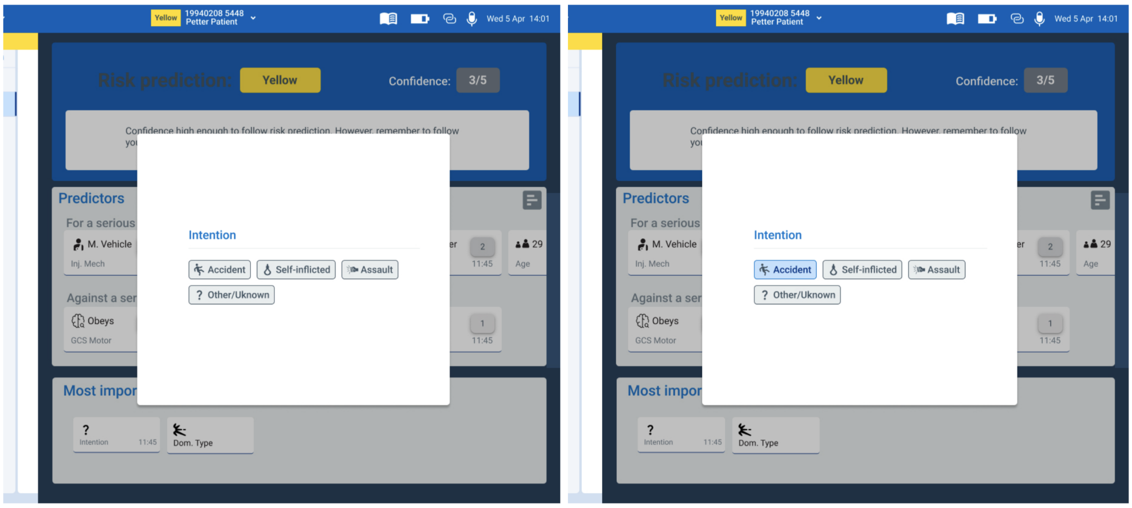


Figure 5.31: Additional variables view of the AI currently lacks but deems important.

it works well for medical professionals. For most parameters, you look at the left Y-axis. It is only Oxygen Saturation (SpO2) and AI confidence value that needs to be checked on the right Y axis. Clicking on the cards on the right should highlight which parameter is selected, but was not functional before testing.

Clicking on the boxes displaying changes in AI confidence should expand as to what the primary reasons for those changes were, however, this was not investigated further due to time and knowledge limitations.

5.6 Test and evaluation

5.6.1 Expert evaluation

An interaction design lecturer with a PhD, an industry UX professional and an interaction design student investigated the system before the final touches were made before testing. The prototype displayed in high fidelity is with minor changes done after the expert evaluation.

The lecturer was positive about the design. The main comment was that some buttons appeared like they could be a bit small for the use case. This was improved for the final prototype. A discussion also took place on whether to use an indicator to show that you could drag out the AI output from the right. Arguments against is that it is a functionality that will be used repeatedly as a professional tool, so after a quick onboarding process and some experience such an indicator would not be needed. It might also create visual excise and not clearly indicate from where on the screen you could drag. It would likely be placed on the edge of the screen while you could drag it from the middle. Arguments for an indicator included that the ambulance service has a lot of personnel rotation, so new users might be common or older users that haven't been active for a while, thus good to have a constant reminder of the functionality. A clear-cut answer was not found during the discussion, thus the indicator was cut to save time.

Figure 5.32: Vital parameters chart

A former UI/UX designer at Avega gave feedback focused on the visual design part. He pinpointed the design mistakes and recommended reconsidering a few things such as; grouping and visual hierarchy, colour and highlight, re-organising focus area, and alignment concerning visual interface design principles, see Figure 5.33.

The student had good things to say. His main gripe was that he wished the checkboxes on the WEST screen were more like buttons. We thought that having them as buttons instead might clutter the interface, so instead created small rounded outlines to help with finding the touch area.

5.6.2 User tests

This usability study was conducted in a setup environment reserve room with a camera and a voice recorder (figure5.34); because conducting field testing in an ambulance is impossible due to the risk of disturbance in the ambulance work routine. The test study used a think-around method while the participant is doing the test and answering follow-up questions at the end of the test.

Testers from VGR were recruited through contacts, mainly using mail. Testers from Halland were recruited at the station between missions. No compensation was advertised for the testing, but they were given chocolate as a thank-you. The table below(Figure 42) displays the distribution of the users, divided by years of experience and the two regions. Note that VGR still uses a paper journaling system for triage, while Halland uses a tablet-based journaling system. There was no difference in

Figure 5.33: Feedback from a UX designer

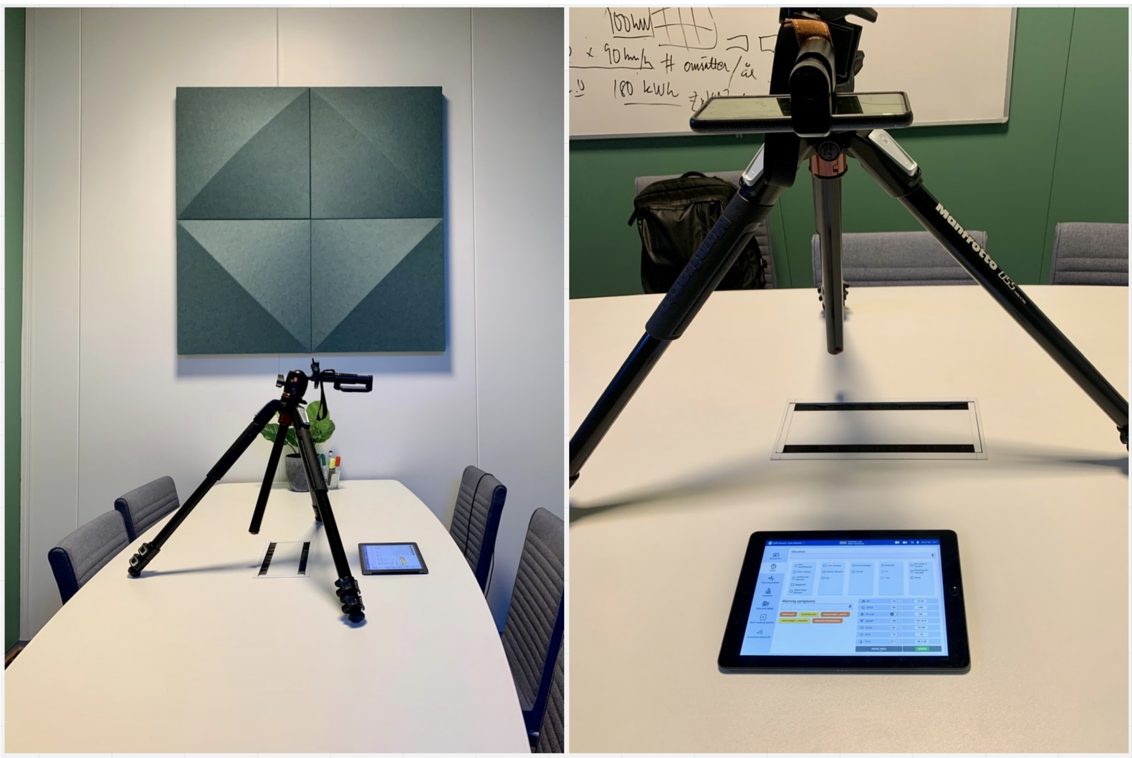


Figure 5.34: User test setting

the testing between the groups. The test took 30-45 minutes. After a 5-minute introduction, each participant performs two major tasks for 15-20 minutes. They then had around 5 to 10 minutes of discussion.

Table 5.1: Distribution of testers

Group	<10 years exp.	>10 years exp.
Halland	1	2
VGR	2	2

According to usability testing standards, the recommended number of participants is 4-8. While 4 is enough to find most errors, having a few more can still help in finding critical problems with the system (Rubin and Chisnell, 2008).

Table 5.2: The table of participants who participated in user testing. Experience shown in years

Subject	Sex	Age	Exp.	Region
no.1	M	43	17	VGR
no.2	M	25	4	VGR
no.3	M	33	10	Halland
no.4	F	47	16	Halland
no.5	F	42	6	Halland
no.6	M	37	6	VGR
no.7	M	48	15	VGR

Figure 5.35 is the setup of the conducted usability test for this study. The introduction was presented as a video recording to maintain the quality of the introduction part the same in each test. The introduction quickly explains the overview of the AI-driven decision support in the design and the overview of the UI layout as we assume that the digital software will need to be trained in the primary use in ambulance education requirements about AI-driven decision support in the future. For the script of the video, see Appendix B.

After the introduction, two prominent medical cases and a task on each were given on paper. Both cases are trauma cases, one a traffic accident of a 29-year-old male and the second an 80-year-old woman falling on a hard floor. Both are based on real cases. The traffic accident is a case that is often over-triaged, and the fall accident is often under-triaged. To read the tasks, see Appendix B. During the test, the participant used a think-aloud protocol (Martin et al., 2012). The participants were asked to speak out their thoughts and what they saw while having their voices and hands recorded. After finishing interacting with the system, a contextual interview about the prototype with a focus on the AI output took place (Figure 5.34).

The results from the usability test are found in the following section. They are divided into different parts, behaviour, which generally means observed interaction

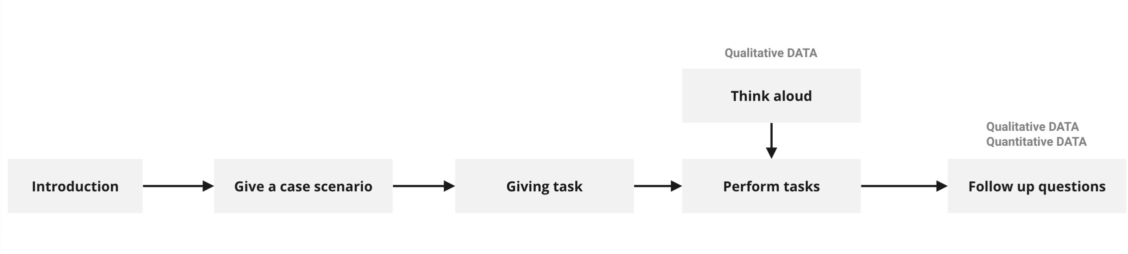


Figure 5.35: Usability testing flowchart

with the system without intervention and discussion, which generally means their reflective thoughts during or after the interaction. A summary is provided, gathering the most important findings.

5.6.2.1 WEST

Behaviour

- Many misses Situation text-field.
- Easier for VGR nurses since Halland nurses don't use WEST.
- Easy for everyone to read and understand the vitals.

Discussion

- Current paper journal is limited in number of preconditions - the digital system could support more and be more specific.

Summary

VGR testers really liked the page, while Halland felt a bit confused since they use a different triaging system. Still, both groups were easily able to interpret and interact with the page. However, the Situation text field took less focus than the rest and was often missed.

5.6.2.2 Scenario

Behaviour

- Many asked for a front and back body.
- Was a bit difficult to be precise with the AIS body, some sections were too narrow.

Discussion

- A future adaptation could be to expand the interface based on what you have put in. For example, if you click on Motor Vehicle you could get more parameters such as seatbelt usage.
- It lacks input for nerve damage outside of the central nervous system.

- One tester thought some Unknown inputs could be removed, as in the case of the GCS motorics.

Summary

It was generally easy to understand the page and fill in the information. The AIS Body input was a bit hard for some, but most had no issues with it. One nurse said they don't use GCS in Halland so that also created some confusion.

5.6.2.3 AI notification and interaction

Behaviour

- Colour-coded notification misinterpreted as AI-led decision.
- Most continued with the interaction on the current page before acting on the notification by opening the AI page.
- First time when opening the AI, about half of the participants opened it from the side and the other half opened it by clicking the top-bar icon. Once this behaviour was set it was usually continued, though some used both.
- AI indicator on the top touch areas was too small and was hard to click for some participants.
- Notification and its text were too small.
- Notification text was not clear enough for new users; users did not understand what to do when the red notification came up saying "Possible high risk. Few parameters required for prediction".

Discussion

- This section was barely mentioned by participants. Some found the notification confusing.

Summary

People did not react directly to the notification. Most noticed it but didn't act on it, others ignored it. Everyone successfully opened the AI page.

5.6.2.4 AI page

Behaviour

- Difficult for many to understand the confidence number on the cards which increases the likelihood of the AI to predict a more serious outcome. Many thought of them as seriousness level on their own, similar to a specific vital sign being Red according to NEWS.
- Many could not discern the difference between the predictors for and against a serious condition, usually because they didn't read the title text.

- Some tried to click on the output cards, which should be intractable. This would allow them to change parameters, view the choices again or get more information. But that wasn't programmed for this prototype.

Discussion

- Predictors and missing variables are good reminders for experienced personnel.
- Both prediction and variables are likely very good for new nurses.
- Majority preferred AI cards to Chart, since it was easy to read and gave a good overview, but Chart was easier to understand for many.
- Easier to take multiple parameters into account, because today a lot of them are in the head.
- In the case of less time-critical patients, some nurses stated that some parameters that might make less sense would be good to have more info on when interacting with them.
- AI cards with missing variables that are interacted with create a sort of back-and-forth communication between the user and the system, making it feel more fluid and increasing trust in the system.
- Should show what kind of situation the AI thinks of, as well as give treatment guidelines.
- Good that the confidence scaling of parameters is not using colours since it might confuse the user.
- Darker shade of grey for stronger prediction was a good help in viewing the page. Helped both in getting an overview and in understanding.

Summary Many had issues with interpreting the different rows of predictors, but after the test with some guidance, most thought it was clear and useful to have both for and against on the same page. Many understood the numbers for confidence on the AI cards, but some misinterpreted them. Halland users had an easier time interacting with the page in general, naturally clicking the missing variables cards. All users found this page helpful, as a reminder for what they could have missed and which predictors were important in the case. AI prediction works like a second opinion, similar to how they currently ask their colleagues about what they might be missing or what triage to make.

5.6.2.5 AI's prediction

Discussion

- Would not trust AI initially if it makes a different decision. However, it might slightly change their own decision, such as if the personnel thinks a patient is most likely yellow but could be orange. An orange or a red AI prediction might change the personnel's decision to be orange.
- AI output can make the personnel rethink their preconceived notions.

- Experienced nurses took in what the AI said, but felt strongly about their own medical decisions and made them themselves, at least when getting used to the AI. Over time, when learning how the system works, it will change when to listen to the system and when not to.
- AI seemed to lack important variables, such as seatbelt usage. Made it harder to trust.
- A participant thought that AI for trauma would be less useful than AI for sepsis, stroke or KOL patients since they don't have the possibility to know what has happened within the body outside the hospital in the case of trauma. The nurse argued that there are a lot of variables that are easy to miss in other cases.
- All participants agreed that the explainability of AI was necessary to trust the system.
- AI confidence could be interpreted differently. One saw confidence three out of 5 as too low, while others mostly trusted that the AI had come to a conclusion based on the guidelines saying that the confidence was high enough, even though they might sometimes disagree with the risk prediction itself.
- Confidence level could help in deciding how much they felt they could trust the AI.
- Despite being given the same information, some completely agreed with the AI decision while others thought the yellow prediction in the traffic case was too low and the red in the fall injury was too high. Often they gave very different reasonings for their thoughts. A partial explanation is likely that they didn't have the patient in front of them, forcing them to fill in more interpretation gaps than usual. Still, some reasoned very differently. In the case of the car accident, some stated red due to the high speeds on the road, while another argued that the kinetics of a side collision is much lower than other kinds of crashes so that the patient could either be green or yellow.

Summary

This is discussed further in the research question discussion.

5.6.2.6 Vital Parameters

Behaviour

- Testers quickly viewed this page when asked after their main tasks had been completed. No one tried to interact with it, but all seemed to quickly be able to interpret it.

Discussion

- Good to keep track of the patient over time. The acuteness of some patients can change a lot.

Summary

Due to not being interactive, it was not tested as much. Most users found it clear and useful.

5.6.2.7 General interface

- Very quick, would help with multiple patients.
- User friendly.
- Experienced Halland nurses already felt they know RETTS by heart, and seldom used the digital version of RETTS since it is too slow, so for them using the system in that way felt new to them. However, one nurse thought that even though they trust their medical judgement, following this system would likely be better for patient safety, reducing biases and errors maintained within automatic thinking. For now, the pad is mostly used to collect vital parameters in Halland before filling out the journal afterwards. Perhaps a more efficient system could help motivate IT system usage.
- While easy to understand and use for most, there should be an educational introduction before using the system.
- Text only was almost always ignored at first. Some read it, but only after first exploring the rest of the interface, which could lead to misunderstanding, in particular when it came to discerning predictors for and against a serious condition in AI output.

Summary

Overall, all users could easily interact with the interface, despite never having used it before. The tasks given were likely a bit unclear for some since they were unsure what to do next in some instances when they had filled in the necessary information on the Scenario page. AI notification seemed to be what caused the most confusion. A reason for this could be that the notification should have been clearer. Another could be that the tasks given during testing might have been too open and should have been stricter about interacting with AI when notified. Finally, having a notification in the middle of filling in information was different.

5.6.2.8 System Usability Scale (SUS)

The SUS was filled in after every session. The results were then collected and calculated to get a SUS score. However, during the analysis of the results, an error in the survey was found. When taking the survey, users were given a question about usability and asked to rate the statement from 1 to 5. In the SUS test, 1 represents Strongly disagree and 5 represents Strongly agree. However, the text describing that was missing when taking the survey. The survey was then most likely interpreted with how much one agrees with the question, rated 1 to 5. Every odd question is positive, and every even question is negative. Both contribute or negate the same amount to the scoring. The scoring is consistent with observation, those who scored the lowest had the most issues interacting with the design, while those that scored highest had almost no issues. Visual analysis indicates generally positive answers, regardless of interpretation. See Appendix D for the distribution of answers for each question.

The SUS score results were, from best to worst: 100, 87.5, 82.5, 77.5, 67.5, 57.5, 52.5.

If the study would have been done correctly, this would give the design an average score of 75, putting it in the top 25% of SUS-tested systems and giving it a “Good” or “B” grading (Lewis, 2018).

No major differences were found between regions, with VGR testers giving it a score of 76.88 and Halland giving it a score of 72.5.

6 Discussion

This chapter starts by answering the research questions. It then discusses the project scope, the research process, and the results presented in the previous chapter. It finishes with our reflections on human factors and ethical perspectives.

6.1 Research questions and answers

RQ1: How can dynamic AI-driven decision support be designed to work well with the ambulance personnel workflow and work environment?

The dynamic approach, allowing AI to give a response at any time while inputting data and actively supporting the user, is intended to speed up the current triage process compared to RETTS or WEST. As soon as the system has enough or close to enough information to make a prediction, it should notify the user. While this makes the decision faster and essential for patient safety, it also obstructs the often-used step-by-step approach of healthcare workers, where they normally make one type of assessment before continuing with the next. The dynamic AI can disrupt these steps at any point, and this creates a change in the workflow. While perhaps not having a negative impact over time, it likely requires training and longer-term evaluation to see how it impacts the workflow.

In addition, the response to the notification was not as expected. Most users ignored it and went on with the journaling process and only went into the AI when they had completely filled in the Scenario page. Whether this was mostly due to instructions, the AI notification not being clear enough, or that it broke with the usual workflow, the reason was not found out during testing. A bigger notification might have helped, but we speculate that training in how to use the AI might be more important. This warrants further investigation, into what the best way to notify the user of a change in AI prediction might be.

Otherwise, the work situation for a nurse can often be stressful and sometimes they need to interact with the system while in a bumpy ambulance. Thus, the information needs to be as clear as possible, and easy to interact with. In this case, the AI cards provided a nice overview and were easy to interact with.

RQ2: How can dynamic AI-driven decision support be designed to work well with the structure of current IT systems used in ambulances?

Throughout this process, we looked at different current ambulance IT systems and found a majority of them currently use a system primarily designed around a tablet-style interface. The navigation was made on the left side of the screen, allowing for switching between different pages. By mirroring this design, it allowed us to explore the possibilities for dynamic AI decision-support integration.

If only designing for a single AI model for trauma, perhaps just another page added to the side would be enough. However, the goal for this system was to think a little bit further ahead and imagine an adaptable system that should be possible to use with other AI models and be able to be accessed at all times. It should be able to respond to the user as soon as enough information is available.

With this in mind, we came up with a design that could be used throughout the entire interface. This allows for continued use of the journaling system, being able to access every page at the same time as the AI is available. Thus, the flow of current systems used in ambulance care would not be disturbed by the proposed overlay design for AI output. Several nurses seem to have appreciated this since it allowed them to remain on the same page and have all the most important variables available at the same time.

While there are likely several other ways to tackle this challenge, we believe we propose a model that could be implemented into current systems without disturbing the current workflow.

RQ3: How important is explainability for ambulance AI-driven decision support?

When asked, all nurses thought of explainability as essential. Just getting a prediction result is not enough. While our sample might not be representative, every single ambulance nurse that tested the system thought it was important to know why the AI came to its conclusion for them to trust it. It was clear that the nurses had high trust in their own medical expertise, and if the system would provide them with a different prediction, it might increase their mistrust of the system. In addition, the explanation given by the AI about how it came to a conclusion gave many benefits that the nurses appreciated that could help in their own evaluation of the patient, such as the most important predictors.

It is important to note that this was of course only the result of a first encounter with the AI. They received a barebones introduction, that it had been validated, was better than previous triage models, and was basing its prediction based on patient registries. This is not a realistic scenario, in real life, they would likely need to receive a much more extensive education. Several nurses mentioned that their trust in the system might increase over time.

While current AI models seem to work well both using transparent and non-transparent algorithms, it is possible that future models that are a lot more accurate could require the latter, in which case the design presented in this project might not be possible. Then proper education and understanding of the system would increase in importance, and we think based on the findings that it would take a longer time for nurses to gain trust.

RQ4: How can AI decision support be explained and visualised to increase trust and understanding of the system?

In this design, the overall confidence level of the AI was described in a box with a number from 1-5. This was easy to interpret for users, but it appeared to be valued differently. While most simply appeared to follow the text saying it was high enough for a risk prediction, others felt that a three out of five was too low in order for them to trust it.

Many had issues interpreting the two rows of predictor cards which were for and against a serious condition. However, just asking users to read the titles or giving a quick explanation was enough for them to feel that the information was clear. The chart many thought made it more intuitive to understand, but the rows of cards were still preferred since they provided an overview that was easier to read.

While briefly investigated during the low-fidelity prototyping phase, this project did not dig deeper into allowing the user to drill down more information about the different variables. This could lead to misunderstandings as to why the AI has interpreted a parameter as particularly serious. In addition, this study did not look into the impact of displaying the combination of two parameters and giving them a combined prediction score.

As an important note for future AI development, nurses reacted negatively when the AI did not take into account measures that they thought would be important. In the case of a traffic injury, the model provided did not take into account seatbelt use or speed on the road. Either AI models need to include such parameters, or there needs to be clear education that these are not important predictors to include in the AI model and that it can make a reliable prediction without them.

RQ5: How could AI-driven decision support impact ambulance workers' cognitive decision process?

A comment that recurred in most tests was that the system presented works as a reminder as to what variables to take into account or record. This was mentioned as a benefit both for inexperienced and experienced nurses. In the latter case, confirmation bias might be a factor which could lead to some predictors not being taken into proper consideration since they don't fit into their initial judgement of the case (Kahneman, 2011).

When it comes to interpreting the final risk prediction, according to the nurses, it can act as a second opinion, similar to that of a colleague. Some thought that if the AI had a different opinion than them, it could force them to rethink, making their decision more conscious and possibly reducing bias. If on the edge between two different risk predictions, some thought they likely would adapt their decision to the AI. Another comment was that if the AI had a similar conclusion, it would make them feel safer in making that judgement.

However, this is only based on the subjective experiences of the nurses' first interactions with the system. This project does not in any way prove that the system makes nurses better at making decisions. For such a conclusion, a bigger sample size and comparative experiments using different systems and measuring the outcome would be required. Still, the findings of this study could work as a foundation for a future study measuring the impact of dynamic AI decision support.

6.2 Project scope

Going into a new field was challenging, with most answers leading to more questions. The healthcare sector in general, and ambulance care in particular is very complex, dealing with regulations, a variety of practices, as many systems and guidelines as many regions, etc. This led to the initial investigatory phases taking longer than anticipated but also granted a strong knowledge base to build upon in the design phase.

While very few XAI projects for ambulance care were found, Andersson and Maxstad (2022) provided a solid foundation for us to work from. This helped speed up the knowledge-acquiring process needed for the extensive field of ambulance care. We felt that it is important to build upon that work, not only developing a new interface for one specific algorithm but to think further about how several types of AI could be integrated into a wider interface. Thus, a focus for this project was to investigate what IT systems are currently available and to make a design that fits into that system. Our interview with Bengt-Arne Sjöqvist also gave us the basis to work towards a more dynamic system that responds to the user in real-time.

Several systems already have satisfactory designs for things like personal information input, what medication has been provided, etc. Thus, the project's focus became to design a way for AI output to be explained, in addition to some form of input to allow the system to be tested in a normal workflow.

The system is designed to be able to work with a variety of different AI models. Bakidou's trauma algorithm was used to make the use cases specific, but it only works with tabular data. Thus, the design should be able to work satisfactorily with AI models using that data type. Other data types, such as image or text analysis were not explored, but even then a similar system could be used where the output of that analysis could hopefully be made into tabular data. An example could be an image of a bruise from a blunt trauma. Let's say that the image analysis deems it as severity 7/10, and that data is provided to the system. However, then the user should be able to drill down on what causes that severity, which we did not explore during user testing.

In general, more investigations on how to drill down to access more information would be needed in future research. Right now, the system does not provide any interactions with the cards in the AI output. They should allow the user to change the values on the cards, but they should also allow the user to gain some more information. For example, in the case of the 80-year-old woman with a spine injury, new nurses might not deem it as serious. They might be surprised by the system deeming it as red, and perhaps not trusting it. There should be some means to get an understanding as to what specific health concern the spine injury provides. A challenge here is that ambulance work is both limited on time and cognitive resources, thus we imagine most of this information would be provided in the Summary page, but exploring this further was outside of the scope of this project.

Finally, the dynamic system currently provided imagines an AI that can consider all kinds of information. Such an AI does currently not exist, but they only consider very specific information about a patient. What information is important for one algorithm might be irrelevant for the next. For example, Bakidou's algorithm doesn't take seat belt usage into account with car accidents, while another algorithm might think of it as very important and put it as a 5 in the contribution to a serious case. Before a dynamic AI that can combine these factors has been developed, perhaps a system to switch between different AI predictions would need to be developed. Perhaps several different AI could work in the background, and the one closest to making a strong or serious prediction could be activated. This is something that needs continued investigation as the development of AI in the medical domain progresses.

6.3 Process

The discovery phase was prolonged within the sixteen-week project, delaying the other phases and making the project quite time constrained. Investigating the field of the healthcare sector is quite a challenge due to its complexity. We had issues with people cancelling or being too busy, and that delayed the definition phase.

Perhaps alternatives to the double diamond model could have helped. The agile approach of the design sprints framework could have led to quicker development of early prototypes that could have let us gather early specific feedback that could have sped up both prototyping and knowledge acquisition. Co-design would have allowed us better access to medical expertise.

However, the design needed to take into account AI development, regulations, the future of healthcare in VGR, current IT solutions, the ambulance work environment, and more. The need to gather information from various sources complicated the knowledge-acquiring process, thus perhaps the prolonged discovery phase was inevitable.

6.4 User study

The ambulance environment can be very stressful and emotional making the ethnography in ambulances somewhat challenging. For example, sitting in the ambulance with EMS personnel could make the patient uncomfortable. First, they needed to understand why the researcher was there and required an explanation. The patient in the second case did not feel like being investigated; she kindly asked the researcher to stand outside the room instead. Moreover, the EMS personnel sometimes asked the researcher for some physical help, which might create bias at some points. Strict regulations made it difficult to document every detail, and photo documentation is only allowed in some situations. The patient cases can differ, so the observation might better take more than one day to gain a deep enough understanding of the ambulance workers and workflows.

The expert interviews were vital in gaining necessary knowledge about the state of the ambulance field and further knowledge about the state of AI in prehospital care. With most, we only had one interview, but it was enough for this project. It was important to gather information from experts from different fields in order to get a full picture.

6.5 Tests and evaluation

Testing using the think-aloud protocol went well, with valuable feedback provided. Questions given to testers were often contextual and based on the interaction. Some were prepared but didn't follow a strict template. This might have made us lose valuable results. For example, AI notification and opening the page could likely have needed further evaluation but were missed when the focus became on the explainability of the AI page.

There are three main issues concerning ecological validity. First, tests were carried out in a laboratory setting. Thus, the vibration felt in an ambulance and the interaction that would entail were not tested for. This mostly would impact the size of texts and buttons, making it harder to read and push the correct buttons. However, the input part of the design is the least important one. The primary goal is to investigate how the interaction and visualisation of AI impact ambulance workers, as well as its integration into a wider system. Thus, the issue of it not being done in an ecological setting is of little concern in this regard.

Second, the scenario tested is not realistic. When it comes to medical tools that impact life and death, they first require extensive education. In this case, the ambulance personnel were given a 3-minute introduction which mostly just introduced the general design and said that the AI was verified. Had they received a full education on the system, it is likely both their thoughts and behaviours would have been different. The primary reason for doing it in this way, outside of lack of resources, time and knowledge to prepare such education, is that we wanted the first thoughts on the system by nurses without being biased beforehand. This stress tests the design, increasing the need for it to be self-explanatory. Thus, if nurses can understand and interact with the design in this scenario, it should be satisfactory for actual use. About half of the nurses did fully understand the role of two rows of predictors initially, and about half only needed a few seconds of explanation to explain it. Thus it appears the design is effective in its communication.

Lastly, one potential improvement would have been to simulate the scenario. An actor could have acted as the patient, and instead of being given a paper with the predictors, the nurse could have received the information required by speaking to the patient. Doing it in this way was an early consideration, and it likely would have increased the validity somewhat. It would also have helped by providing more context for the nurses, making it easier to imagine using it in a real scenario. It would likely be more resource and time-consuming, however, and it's unclear how big the benefits would be.

Due to the error in SUS, it is difficult to say how much it can be interpreted. We argue it can likely be seen as a good indicator of usability since the statements are the same, and the odd questions are positive while the even questions are negative, thus the results likely even out, and the score can likely still be used as a good indicator. Even without taking SUS into account, observation showed positive results whereas most actions were successful without assistance. This, despite the design being done with long-term use in mind, and not intended to be used without training. We thus argue that the design holds a high level of usability. It should be useful in ambulance care if fully developed into a working IT system.

6.6 Human factors

No proper testing was conducted to test the cognitive effects of the design. For example, for cognitive load, other tasks would be required to be done simultaneously and to avoid bias it would require a bigger sample and something to compare it to. However, the system was designed with these factors in mind and the following is an analysis of possible effects.

6.6.1 Wickens attentional processing model

Several parts of the design tried to limit the number of required actions, which should limit the required attentional resources. For example, trauma scenarios and injury mechanics were auto-selected, meaning almost all steps in Wicken's model can be skipped.

Within the AI output, the Most important missing variables make the search for these variables much easier, allowing for a more automatic approach, and requiring fewer resources. However, ambulance personnel often work on a step-by-step approach, first doing an assessment of the most important vital signs, before going to vital parameters, looking at preconditions, etc. This approach, of jumping to different most important parameters might make the work conditions less structured, which instead might require more attentional resources. Perhaps the AI notification might also obstruct the workflow. In these cases, we also haven't tested how this would work in a situation with a patient. Most of the input used is mostly observational and does not require too much questioning of the patient. In a case when a patient is being asked questions, perhaps the variables asked for can appear very random for the patient, making them more difficult to answer.

6.6.2 4-D multiple resource model

The system itself primarily uses visio-spatial perception resources. In many cases, symbols are used, which should reduce Visio-verbal load, such as reading text. This frees up verbal-auditory encoding, making it easier for the personnel to listen to a patient while using the system. However, the system is still text-heavy, especially when it comes to pre-conditions and warning symptoms.

Additionally, the system is visio-spatial heavy, which might lead to the user regularly missing important parts of the interface. Colours, contrast and movement have been used to guide attention, but the spatial-auditory perception channel remains unused. This is an area for further exploration if sounds could be used when the AI changes prediction. Of course, the changing work environment with ever-changing sounds might make this difficult. Haptic vibrations, similar to those used on phones, could also help with notifying the user. But there is also a balance on how attention intrusive the system should be, to not interrupt the workflow.

6.6.3 Bias

Bias was a serious consideration during the design. For example, AI output cards currently use different grey levels combined with numbers to indicate confidence contribution for each parameter. This includes both variables making a worse outcome more likely and variables indicating the opposite. An alternative to discern the two could have been different colours, such as red for a worse outcome and green for better outcome. However, red is also used to indicate a very serious time-dependent case, and thus a variable with red as a colour might make it appear more serious than it is, such as in a yellow prediction patient case, as an example.

Overall, the design should help against bias. Some variables might seem less important for an ambulance worker, or some symptoms might be ignored. It is impossible for a human to know the importance of every variable, forcing shortcuts which enable bias. The list of most important predictors in the AI output should bring these variables to the forefront, forcing the user to rethink and reevaluate variables that might otherwise have been ignored. The same goes for when the AI comes to a different conclusion, forcing the ambulance worker to rethink. However, if the AI would be just as incorrect as the personnel, there is a risk of enhancing the decision based on bias rather than suggesting against it. Of course, the precision of the algorithm is what counts here.

6.6.4 Gestalt principles

These design principles were taken into consideration in many cases. Law of common region is used, especially with the variables in Scenario, making it easy to group the objects together.

However, the principles can also explain one of the key problems of the design. On the AI output page, many testers had problems discerning the two kinds of predictors, as well as missing variables. Despite them being grouped very differently with different regions and proximity, they appeared to be too similar for initial users. This was partially by design, they are similar in that they are all displaying variables of a similar kind, but they display slightly different things. Perhaps colour here would make a difference, but that might increase the risk of bias.

6.7 Ethical aspects

When it comes to the ethics of the process, privacy was taken into account during the ethnography phase, and all information that could be linked to a patient has been censored. We tried to make our interactions with medical professionals as efficient as possible to respect their time and energy.

About the design itself, based on feedback from users and experts, the design has a high potential to improve the workflow of nurses, especially those with less experience or who are rusty from taking a few years' breaks from intensive work as an ambulance worker. Thus, the design has the potential to reduce stress and increase patient safety.

Although slightly outside the scope, this project touches upon when an AI should be allowed to make a decision. As it is currently imagined, the AI should only make a recommendation to the medical professionals, it is still a person responsible for the final decision. At what confidence level and at what acuteness level the AI should make the recommendation is up for regulators to decide. Here we have imagined a scenario where acuteness level red requires a lower confidence level than yellow before giving out a recommendation. AI is also only as good as its data, with biassed data creating biassed outcomes (Zou & Schiebinger, 2018). This is however more a question for developers of the algorithm than it is for interaction design.

Some questions remain, at what level will it actually be the nurse making the decision and not the AI? Could the nurse become over-reliant on AI in making the decisions? Hagiwara et. al., (2013) shows that computerised decision support increases compliance with regional guidelines compared to paper. It might appear likely then that AI could increase compliance even further. This might be a good thing since AI is hopefully better at making decisions than current systems. But what happens when someone is seriously hurt because the AI makes the wrong decision, deeming the patient as less serious?

People seem to prefer humans making decisions instead of AI (Leyer, & Schneider, 2019). One could then argue that these kinds of decisions are already being made with existing decision support systems. An AI is mostly a tool to accomplish the same thing but be more precise. It remains important for the nurse to not become over-reliant on the AI, especially in the cases when the AI appears to make a decision that points to a less severe outcome. Good training on how the system works and knowledge about biases such as confirmation bias (Kahneman, 2011) are necessary. We try in the design to remind the user to use their expertise. The design would have to be evaluated for long-term use to know whether this is enough or if the design needs to be improved further to empower the medical professionals to make their own decisions.

7 Conclusion

This study proposes a design for XAI in ambulances that should work well overall with the workflow of ambulance personnel and does not disrupt the structure of current IT systems. It allows nurses to constantly receive feedback from the system, which hopefully would enable them to make quicker and better decisions.

However, the design was not tested long-term and only used subjective measures. Thus, how ambulance personnel would adapt to the system over time or whether it actually would allow them to make better decisions was not studied. Nevertheless, the feedback received from nurses was overall positive and provided promising results for future studies.

We have four main recommendations for future studies:

1. Comparative quantitative studies of triage results, with or without XAI.
2. Testing displaying the risk prediction of a combination of predictors instead of only showing it for single predictors.
3. Testing the impact and usability of drilling down more detailed information about the predictors.
4. Testing out alternatives to our notification solution to help guide the user when to use the system.

We believe that AI has a big potential for the future of ambulance care. A current challenge is that the IT systems are rarely used by experienced EMS personnel, thus the design must provide enough benefits while also being very effective in usability and speed to promote usage. Regulations will also need to guide how confident the AI needs to be in order to produce different predictions.

8

References

Antoniadi, A. M., Du, Y., Guendouz, Y., Wei, L., Mazo, C., Becker, B. A., & Mooney, C. (2021). Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: a systematic review. Applied Sciences, 11(11), 5088.

Andersson, J., & Maxstad, E. (2022). User Interface concept for AI-supported clinical decision making in prehospital emergency care.

Andersson J., Maxstad E.(2022).User Interface concept for AI-supported clinical decision making in prehospital emergency care: A study & development project in how to implement new technology for Emergency Medical Services nurses

Abràmoff, M. D., Lavin, P. T., Birch, M., Shah, N., & Folk, J. C. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. npj Digit. Med. (2018).

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. Intl. Journal of Human–Computer Interaction, 24(6), 574-594.

Bakidou, A., Caragounis, E-C., Andersson Hagiwara, M., Jonsson, A., Sjöqvist, B.A., Candefjord, S., "On Scene Injury Severity Prediction (OSISP) Model for Trauma Developed Using the Swedish Trauma Registry," 2023. Submitted

Bieler, D., Kollig, E., Hackenberg, L., Rathjen, J. H., Lefering, R., & Franke, A. (2021). Penetrating injuries in Germany–epidemiology, management and outcome an analysis the TraumaRegister DGU®. Scandinavian journal of trauma, resuscitation and emergency medicine, 29(1), 1-14.

Blix, M., & Levay, C. (2018). Operation digitalisering: en ESO-rapport om hälso- och sjukvården. Regeringskansliet, Finansdepartementet. British Design Council.(2019). Framework for Innovation: Design Council's evolved Double Diamond. https://www.designcouncil.org.uk/our-work/skills-learning/tools- frameworks/framework-for-innovation-design-councils-evolved-double-diamond/

Buendia, R., Candefjord, S., Fagerlind, H., Bálint, A., & Sjöqvist, B. A. (2015). On scene injury severity prediction (OSISP) algorithm for car occupants. Accident Analysis & Prevention, 81, 211-217.

Bresciani, S., & Porta, M. (2012). Evaluating the Usability and Learnability of Three Icon Sets. In CHItaly '12: Proceedings of the 9th ACM Conference on Creativity &

Cognition (pp. 348-349). Association for Computing Machinery.

Bernard, H. R. (2017). Research methods in anthropology: Qualitative and quantitative approaches. Rowman & Littlefield.

Boote, D. N., & Beile, P. (2005). Scholars before researchers: On the centrality of the dissertation literature review in research preparation. Educational Researcher, 34(6), 3-15. Hart, C. (2018). Doing a literature review: Releasing the social science research imagination. Sage.

Buxton, B. (2007). Sketching User Experiences: Getting the Design Right and the Right Design. Morgan Kaufmann.

Buchanan, R. (1992). Wicked problems in design thinking. Design Issues, 8(2), 5-21.

Candefjord, S., Muhammad, A. S., Bangalore, P., & Buendia, R. (2021). On Scene Injury Severity Prediction (OSISP) machine learning algorithms for motor vehicle crash occupants in the US. Journal of Transport & Health, 22, 101124.

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. Journal of the American statistical association, 79(387), 531-554.

Cheema B, Twomey M. In: Health Do, editor. The South Africa Triage Scale (SATS), Training Manual 2012: Emergency Medicine Society of South Africa. Cape Town: Western Cape Government Department of Health; 2012. https://emssa.org.za/wp-content/uploads/2011/04/SATS-Manual-A5-LR-spreads.pdf.

Chittaro, L. (2010). Designing mobile augmented reality applications. Springer.

Cross, N. (2006). Designerly ways of knowing. Springer.

Cooper, A., Reimann, R., & Cronin, D. (2014). About face 3: The essentials of interaction design. John Wiley & Sons.

Dana, C.,Jeffrey, R.(2008).Handbook of usability testing – How to Plan, Design and Conduct Effective Tests. Second Edition. Hoboken : John Wiley & Sons, Incorporated, 2008.

Dejan Todorovic (2008) Gestalt principles. Scholarpedia, 3(12):5345.

Design Council. (2005). The double diamond: Design process. https://www.designcouncil.org.uk/news-opinion/double-diamond-design-process

Design Council. (2015). A guide to the double diamond. https://www.designcouncil.org.uk/news-opinion/guide-double-diamond

Ethics and Governance of Artificial Intelligence for Health: WHO Guidance. World Health Organization, Geneva. <https://www.who.int/publications/i/item/9789240029200>

El-Sheikh, E., Baker, T., & Williams, M. (2015). Using weighted decision matrices to support engineering design. The Design Society.

Eurostat. (2022). Eurostat Statistics explained. Eurostat Statistics Explained.

https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Population structure and ageing:The share of elderly people continues to increase

Fraser HSF, Cohan G, Koehler C, Anderson J, Lawrence A, Pateña J, Bacher I, Ranney ML. Evaluation of Diagnostic and Triage Accuracy and Usability of a Symptom Checker in an Emergency Department: Observational Study. JMIR Mhealth Uhealth. 2022 Sep 19;10(9):e38364. doi: 10.2196/38364. PMID: 36121688; PMCID: PMC9531004.

Fernandes, M., Vieira, S. M., Leite, F., Palos, C., Finkelstein, S., & Sousa, J. M. (2020). Clinical decision support systems for triage in the emergency department using intelligent systems: a review. Artificial Intelligence in Medicine, 102, 101762.

Garrett, J. J. (2011). The elements of user experience: User-centered design for the web and beyond. New Riders.

Gasson, S., & Kivits, J. (2018). The use and effectiveness of the double diamond design framework in healthcare innovations: A systematic review. BMC Health Services Research, 18(1), 860. https://doi.org/10.1186/s12913-018-3662-y

Goodwin, K. (2009). Designing for the Digital Age: How to Create Human-Centered Products and Services. Wiley.

Google.(2023). Design Sprint Kit. URL: https://designsprintkit.withgoogle.com/

Google. (n.d.). Chips – material design 3. Material Design. https://m3.material.io/components/chips/
Region Stockholm. (2021). Medicinska Behandlingsriktlinjer för ambulanssjukvården.https://vardgivar
vard/ambulanssjukvard/medicinska behandlingsriktlinjer-for-ambulanssjukvarden.pdf

Goldenberg, S. L., Nir, G., & Salcudean, S. E. (2019). A new era: artificial intelligence and machine learning in prostate cancer. Nature Reviews Urology, 16(7), 391-403.

Greenspan, L., McLellan, B. A., & Greig, H. (1985). Abbreviated injury scale and injury severity score: A scoring chart. The Journal of trauma, 25(1), 60-64.

Hagiwara, M. A., Sjöqvist, B. A., Lundberg, L., Suserud, B. O., Henricson, M., & Jonsson, A. (2013). Decision support system in prehospital care: a randomized controlled simulation study. The American journal of emergency medicine, 31(1), 145-153.

Hanington, B., Martin, B.(2012).Universal Methods of Design : 100 Ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions.Quarto Publishing Group USA

Hassenzahl, M., & Tractinsky, N. (2006). User experience - a research agenda. Behaviour & Information Technology, 25(2), 91-97.

Hassan, H., Ren, Z., Zhao, H., Huang, S., Li, D., Xiang, S., & Huang, B. (2022). Review and classification of AI-enabled COVID-19 CT imaging models based on computer vision tasks. Computers in biology and medicine, 141, 105123.

Hammersley, M., & Atkinson, P. (2007). Ethnography: Principles in practice (3rd ed.). Routledge.

HSLF-FS 2018:38.The Health and Medical Care Act (HSL), Chapter 5, 1§

Holtzblatt, K., & Beyer, H. (2017). Contextual design: Design for life. Morgan Kaufmann.

Houde, S., & Hill, C. (1997). What do prototypes prototype?. In Handbook of human-computer interaction (pp. 367-381). North-Holland.

He, H., Gray, J., Cangelosi, A., Meng, Q., McGinnity, T.M., Mehnen, J.(2022). The Challenges and Opportunities of Human-Centered AI for Trustworthy Robots and Autonomous Systems. IEEE. DOI: 10.1109/TCDS.2021.3132282

Interaction Design Foundation. (2002). User Experience: Beginner guide, What is User Experience?. URL: https://www.interaction-design.org/courses/user-experience-the-beginner-s-guide-deprecated-v2/lessons/1.2

ISO 9241-210 (2019). Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems. CEN-CENELEC Management Centre: Avenue Marnix 17, B-1000 Brussels.

Imrie, R., & Hall, P. (2003). Inclusive design: designing and developing accessible environments. Taylor & Francis.

Jesse James Garrett, The Elements of User Experience: User-Centered Design for the Web and Beyond (2nd Edition), New Riders, 2010

Jonsson, A., Bohm, K., Castrèn, M., & Svensson, L. (2020). Decision support system improves prehospital care. Prehospital Emergency Care, 24(2), 203-213.

Kim Y, Groombridge C, Romero L, Clare S, Fitzgerald MC. Decision Support Capabilities of Telemedicine in Emergency Prehospital Care: Systematic Review. J Med Internet Res. 2020 Dec 8;22(12):e18959. doi: 10.2196/18959. PMID: 33289672; PMCID: PMC7755537.

Karolinska Institution(2020). Reference guide for APA 7. https://kib.ki.se/en/write-cite/writing-references-apa-vancouver/reference-guides/reference-guide-apa-7

Kahneman, D. (2011). Thinking fast and slow. Macmillan.

Kale, P. (2020). User journey maps: What they are and how to create one. InVision. Retrieved from https://www.invisionapp.com/inside-design/user-journey-maps/

Keselman, B., Berglund, A., Ahmed, N., Grannas, D., von Euler, M., Holmin, S. & Mazya, M. V. (2022). Analysis and modelling of mistriage in the Stockholm stroke triage system. European Stroke Journal, 7(2), 126-133.

Kumar, A., & Chakrabarti, A. (2007). A weighted matrix approach to product positioning: A case study from the mobile phone market. Journal of Engineering Design, 18(4), 355-368.

Leyer, M., & Schneider, S. (2019). Me, you or AI? How do we feel about delegation.

Lewis, J. R. (2018). The system usability scale: past, present, and future. International Journal of Human–Computer Interaction, 34(7), 577-590.

Luschekina, A. (2019). Using customer journey mapping in service design: A practical guide. Service Design Network. Retrieved from https://www.service-design-network.org/community-knowledge/using-customer-journey-mapping-in-service-design-a-practical-guide

Martin, B., Hanington, B., & Hanington, B. M. (2012). Universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions. Rockport Pub.

Material Design. (2023). Material 3: the latest version of Google's open-source design system. Design and build beautiful, usable products with Material 3. https://m3.material.io/

Magnusson, C. (2021) Patient Assessment and Triage in Emergency Medical Services. The Swedish EMS nurse in a new role. [Doctoral dissertation, University of Gothenburg. Sahlgrenska Academy]. URL: https://gupea.ub.gu.se/handle/2077/67134

Majrashi, K. (2022). Performance of mobile users with text-only and text-and-icon menus in seated and walking situations. Behaviour & Information Technology, 41(1), 32-50.

Islam, M. N. (2015, December). Exploring the intuitiveness of iconic, textual and icon with texts signs for designing user-intuitive web interfaces. In 2015 18th International Conference on Computer and Information Technology (ICCIT) (pp. 450-455). IEEE.

Magnusson, C., Herlitz, J., Karlsson, T., Jiménez-Herrera, M., & Axelsson, C. (2019). The performance of the EMS triage (RETTS-p) and the agreement between the field assessment and final hospital diagnosis: a prospective observational study among children< 16 years. BMC pediatrics, 19(1), 1-10.

Mehrotra, A., & Hendrix, D. (2019). Interaction design and artificial intelligence: Challenges and opportunities. In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1-8).

Meroni, A., Sangiorgi, D., & Prendiville, A. (2019). Design for service innovation and development. Routledge.

Moussavi, M., Szymkowiak, A., & Lee, J. (2019). Customer journey maps in experience design. Journal of Design, Business & Society, 5(1), 1-16. doi: 10.1386/$dbs_{0}0007_{1}$

National Early Warning Score (NEWS) 2 Standardising the assessment of acute-illness severity in the NHS. Royal college of physicians, 2017

Magnusson, C. (2021). Patient Assessment and Triage in Emergency Medical Services-The Swedish EMS nurse in a new role.

Nazir T, Mushhood Ur Rehman M, Asghar MR, Kalia JS. Artificial intelligence assisted acute patient journey. Front Artif Intell. 2022 Oct 4;5:962165. doi: 10.3389/frai.2022.962165. PMID: 36267660; PMCID: PMC9577284.

Norman, D. A. (2002). The design of everyday things. Basic books.

Norman, D. A., & Nielsen, J. (2010). The definition of user experience. Retrieved from https://www.nngroup.com/articles/definition-user-experience/

Norman, D. A. (2013). The design of everyday things: Revised and expanded edition. Basic books.

National Health Service (NHS). (2023). Design and build digital services for the NHS. https://service-manual.nhs.uk/design-system/design-principles

O'Brien, J., & Rodden, T. (2017). Human-centred design. Interactions, 24(1), 26-31. Vredenburg, K., Isensee, S., & Righi, C. (2002). User-centred design: An integrated approach. Prentice Hall.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). The measurement of meaning. University of Illinois Press.

Porter A, Badshah A, Black S, Fitzpatrick D, Harris-Mayes R, Islam S, Jones M, Kingston M, LaFlamme-Williams Y, Mason S, McNee K, Morgan H, Morrison Z, Mountain P, Potts H, Rees N, Shaw D, Siriwardena N, Snooks H, Spaight R, Williams V. Electronic health records in ambulances: the ERA multiple-methods study. Southampton (UK): NIHR Journals Library; 2020 Feb. PMID: 32119231.

Prop. 1996/97:60. Prioriteringar inom hälso- och sjukvården. Retrieved from *https : //www.riksdagen.se/sv/dokument−lagar/dokument/proposition/prioriteringar− inom − halso − −och − sjukvarden*$_G$*K*0360

Razzak, M. I., Naz, S., & Zaib, A. (2018). Deep learning for medical image processing: Overview, challenges and the future. Classification in BioApps: Automation of Decision Making, 323-350.

Reynolds, G. (2018). The sketchbook challenge: Techniques, prompts, and inspiration for achieving your creative goals. Watson-Guptill.

Rubin, J., & Chisnell, D. (2008). Handbook of usability testing: How to plan, design, and conduct effective tests. Wiley.

Russell, S. J., & Norvig, P. (2010). Artificial intelligence: A modern approach. Prentice Hall.

Sanders, E. & Stappers, P.J. (2008): Co-creation and the New Landscapes of Design, in Journal of CoDesign, 4:1, pp. 5-18

SALAR (Swedish Association of Local Authorities and Regions). (2019). National strategy for eHealth in prehospital care. Retrieved from https://www.government.se/contentassets/65d ehealth—the-strategy-for-accessible-and-secure-information-in-health-and-social-care-s2011.023

SCB. (2022). SCB publicerar Stor Kartläggning av Sveriges äldre. SCB. https://www.scb.se/pressmedd publicerar-stor-kartlaggning-av-sveriges-aldre/

Stickdorn, M., & Schneider, J. (2012). This is service design thinking: Basics, tools, cases. Amsterdam: BIS Publishers.

SFS 2017:30 Hälso- och Sjukvårdslagen. Retrieved from *https : //www.riksdagen.se/sv/dokument− lagar/dokument/svensk−forfattningssamling/halso−−och−sjukvardslag*$_s$*fs*− 2017 − 30

Snyder, C. (2003). Paper Prototyping: The Fast and Easy Way to Design and Refine User Interfaces. Morgan Kaufmann.

Sharp H., Preece J., Rogers Y. (2019). Interaction Design: Beyond Human-Computer Interaction. John Wiley & Sons, Incorporated

SOSFS 2009:10. Socialstyrelsens föreskrifter om ambulanssjukvård m.m., Chapter 2, 1§. Retrieved from $https : //www.socialstyrelsen.se/kunskapsstod - och - regler/regler - och - riktlinjer/foreskrifter - och - allmanna - rad/konsoliderade - foreskrifter/200910 - om - ambulanssjukvard - m.m/$

Socialstyrelsen(2023). Sveriges prehospitala Akutsjukvård - nulägesbild, bedömning och utvecklingsförslag. Retrieved from https://www.socialstyrelsen.se/om-socialstyrelsen/pressrum/pre utmaningar-for-ambulanssjukvarden/

state-of-the-art adjective - Definition, pictures, pronunciation and usage notes | Oxford Advanced Learner's Dictionary at OxfordLearnersDictionaries.com. (n.d.). https://www.oxfordlearnersdictionaries.com/definition/english/state-of-the-art

Sveriges läkarförbund. (2021, August 31). Digital Utveckling I vården. Sveriges läkarförbund. https://slf.se/var-politik/digital-utveckling-i-varden/

Teasdale, G., Maas, A., Lecky, F., Manley, G., Stocchetti, N., & Murray, G. (2014). The Glasgow Coma Scale at 40 years: standing the test of time. The Lancet Neurology, 13(8), 844-854.

Tran, J., & Winkler, J. (2021). Sjuksköterskors upplevelser av ett nytt triagesystem på en akutmottagning i förhållande till patientsäkerhet.

Tsekleves, E., Gonzalez, R., Kay, J., & Harrison, D. (2020). A systematic review of research on the Double Diamond design framework in the last decade. International Journal of Design, 14(1), 1-20. https://doi.org/10.1186/s12913-020-05560-5

Vicente, V., Johansson, A., Selling, M. et al. Experience of using video support by prehospital emergency care physician in ambulance care - an interview study with prehospital emergency nurses in Sweden. BMC Emerg Med 21, 44 (2021). https://doi.org/10.1186/s12873-021-00435-1

Ware, C. (2019). Information visualization: perception for design. Morgan Kaufmann.
Wickens, C. D., Helton, W. S., Hollands, J. G., & Banbury, S. (2021). Engineering psychology and human performance - Chapter 1. Routledge.

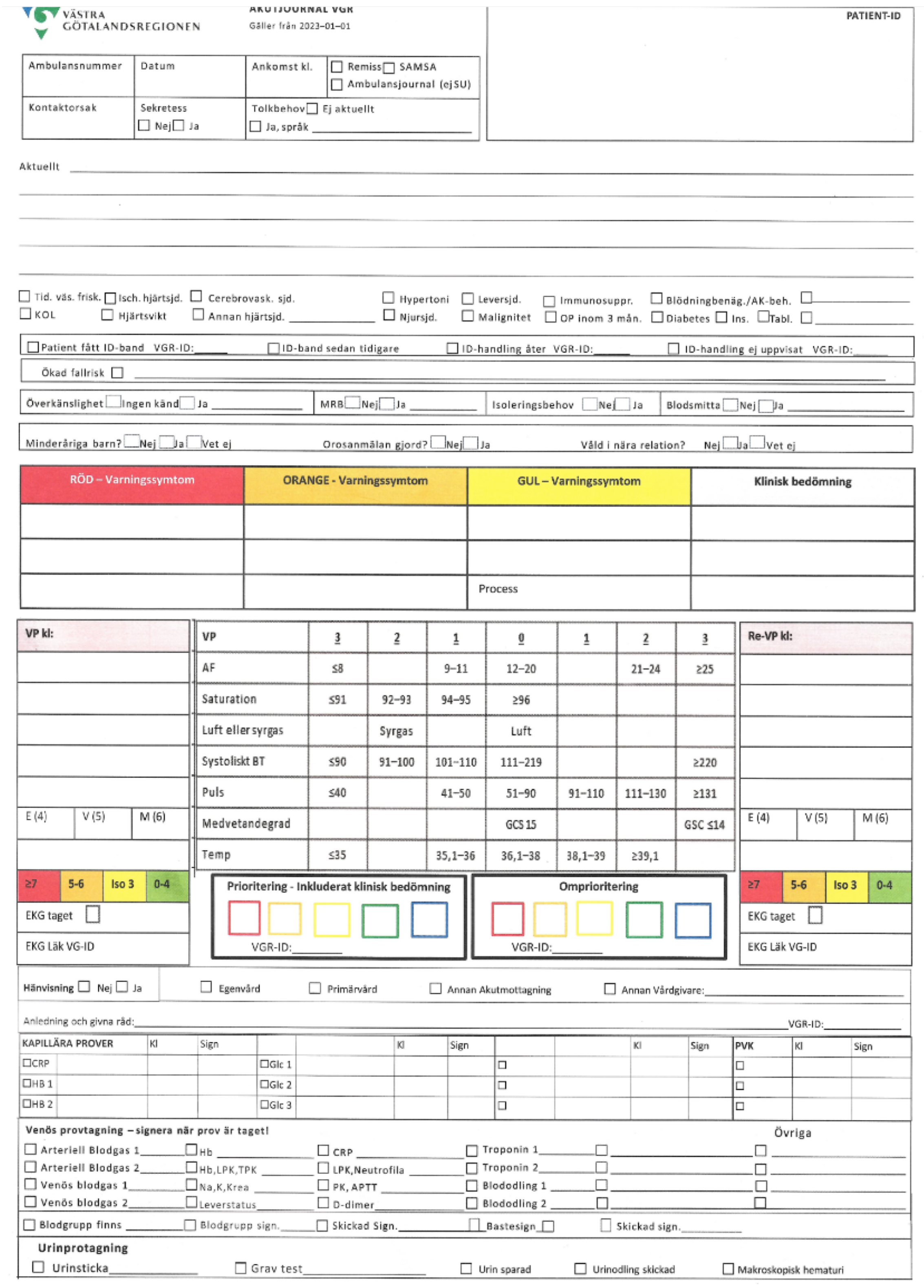
Wickens, C. D. (2008). Multiple resources and mental workload. Human factors, 50(3), 449-455.

Zou, J., & Schiebinger, L. (2018). AI can be sexist and racist—it's time to make it fair.

Zuiderveen Borgesius, F. (2018). Discrimination, artificial intelligence, and algorithmic decision-making. línea], Council of Europe.

A

VGR paper journal



VÄSTRA GÖTALANDSREGIONEN

AKUTJOURNAL VGR
Gäller från 2023–01–01

PATIENT-ID

Ambulansnummer	Datum	Ankomst kl.	☐ Remiss ☐ SAMSA ☐ Ambulansjournal (ej SU)
Kontaktorsak	Sekretess ☐ Nej ☐ Ja	Tolkbehov ☐ Ej aktuellt ☐ Ja, språk ___	

Aktuellt ___

☐ Tid. väs. frisk. ☐ Isch. hjärtsjd. ☐ Cerebrovask. sjd. ☐ Hypertoni ☐ Leversjd. ☐ Immunosuppr. ☐ Blödningbenäg./AK-beh. ☐ ___
☐ KOL ☐ Hjärtsvikt ☐ Annan hjärtsjd. ___ ☐ Njursjd. ☐ Malignitet ☐ OP inom 3 mån. ☐ Diabetes ☐ Ins. ☐ Tabl. ☐ ___

☐ Patient fått ID-band VGR-ID: ___ ☐ ID-band sedan tidigare ☐ ID-handling åter VGR-ID: ___ ☐ ID-handling ej uppvisat VGR-ID: ___

Ökad fallrisk ☐ ___

Överkänslighet ☐ Ingen känd ☐ Ja ___ | MRB ☐ Nej ☐ Ja ___ | Isoleringsbehov ☐ Nej ☐ Ja | Blodsmitta ☐ Nej ☐ Ja ___

Minderåriga barn? ☐ Nej ☐ Ja ☐ Vet ej | Orosanmälan gjord? ☐ Nej ☐ Ja | Våld i nära relation? Nej ☐ Ja ☐ Vet ej

RÖD – Varningssymtom	ORANGE - Varningssymtom	GUL – Varningssymtom	Klinisk bedömning
		Process	

VP kl:
E (4) / V (5) / M (6)
≥7 / 5-6 / Iso 3 / 0-4
EKG taget ☐
EKG Läk VG-ID

VP	3	2	1	0	1	2	3
AF	≤8		9–11	12–20		21–24	≥25
Saturation	≤91	92–93	94–95	≥96			
Luft eller syrgas		Syrgas		Luft			
Systoliskt BT	≤90	91–100	101–110	111–219			≥220
Puls	≤40		41–50	51–90	91–110	111–130	≥131
Medvetandegrad				GCS 15			GSC ≤14
Temp	≤35		35,1–36	36,1–38	38,1–39	≥39,1	

Re-VP kl:
E (4) / V (5) / M (6)
≥7 / 5-6 / Iso 3 / 0-4
EKG taget ☐
EKG Läk VG-ID

Prioritering - Inkluderat klinisk bedömning ☐ ☐ ☐ ☐ ☐ VGR-ID: ___

Omprioritering ☐ ☐ ☐ ☐ ☐ VGR-ID: ___

Hänvisning ☐ Nej ☐ Ja ☐ Egenvård ☐ Primärvård ☐ Annan Akutmottagning ☐ Annan Vårdgivare: ___

Anledning och givna råd: ___ VGR-ID: ___

KAPILLÄRA PROVER	Kl	Sign		Kl	Sign		Kl	Sign	PVK	Kl	Sign
☐CRP			☐Glc 1			☐			☐		
☐HB 1			☐Glc 2			☐			☐		
☐HB 2			☐Glc 3			☐			☐		

Venös provtagning – signera när prov är taget! | Övriga

☐ Arteriell Blodgas 1___ ☐ Hb ___ ☐ CRP ___ ☐ Troponin 1___ ☐ ___ ☐ ___
☐ Arteriell Blodgas 2___ ☐ Hb,LPK,TPK ___ ☐ LPK,Neutrofila ___ ☐ Troponin 2___ ☐ ___ ☐ ___
☐ Venös blodgas 1___ ☐ Na,K,Krea ___ ☐ PK, APTT ___ ☐ Blododling 1 ___ ☐ ___ ☐ ___
☐ Venös blodgas 2___ ☐ Leverstatus ___ ☐ D-dimer ___ ☐ Blododling 2 ___ ☐ ___ ☐ ___

☐ Blodgrupp finns ___ ☐ Blodgrupp sign.___ ☐ Skickad Sign.___ ☐ Bastesign ☐ ☐ Skickad sign.___

Urinprotagning
☐ Urinsticka___ ☐ Grav test___ ☐ Urin sparad ☐ Urinodling skickad ☐ Makroskopisk hematuri

Figure A.1: VGR paper journal

B

Video introduction and tasks for testing

B.1 Script for video introduction

Imagine yourself 8 years into the future. You are about to use a new ambulance IT-system for the first time. It has AI decision support integrated into the system.

The AI triage system has been validated as a more precise triage system then previous systems, such as RETTS or WEST. However, it still needs to be tested with ambulance personnel, how they react to it and trust it. For this reason, WEST is still part of the system so that it can be compared to the AI triage system.

While WEST is still there to help you, you are now expected to primarily use the AI, of course combined with your own medical expertise to make a decision about the patient.

The AI takes into account all the information you put into the system. Then it makes a prediction based on that information. It has been trained on patient databases where it compares different predictors impacting the patient to the patient outcomes.

Here is a brief overview of the system. To the left you have the main navigation. For this test you can only access WEST and Scenario pages since they are the primary pages for triage. Vital parameters for NEWS are automatically updated in the system by using a monitoring device on the patient.

In the top you have an indicator for the AI, showing you what prediction the AI is currently doing. To get more details about how the AI came up with their prediction, you can click this indicator. You can also slide from any part of the page from the right, and drag it out. To remove the output, simply drag it back or click outside of it.

While inputting information into the system, at any point you might get a notification from the system that the AI has changed or need more information. It would be good at this point to see what caused this, and make a triage dependent on the information you get from the AI. The AI might also ask for more input. Feel free to follow that if you think you should.

Please note that this early prototype is a simulation of the actual system. Thus, it can act unexpectedly. For example, clicking on the Situation text bar automatically inputs text instead of using a keyboard or dictation.

In order for us to be able to interpret what you do during the test, we ask you to speak your thoughts during the test. For example. it could be that when you see the navigation bar on the left, that you spontaneously think it looks clear and good, or that you get confused and don't know how to use it.

B.2 Tasks

During this test you are instructed to do different tasks while thinking-aloud. You are to verbalize what you think as you perceive and act on the system. If you want to point to where you have your focus with your finger, this can help too.

You will now be granted two completely separate cases.

Case 1
Male, 29, car accident, 100km/h road. Side-to-side collision, truck hit the patient's car, passenger car into guardrail, rotation of passenger car to a stop. You record his vitals through a device that is connected to the IT interface and automatically updates it. The driver/passenger has been wearing a seat belt. Slight blunt injury to arms/upper extremities.

Tasks:
You will be granted 2 main tasks. At any point, the decision support system might give you a notification.

When given a notification, look through the current information on the decision support page. Consider making a decision based on the data provided, or filling in important missing variables. If the latter, start with the most important variable.

1. On the **WEST** page, look through the vitals and check that they are ok. Fill in **Situation** information and **Warning symptoms**, in this case moderate pain.

2. Go to the **Scenario** page, and fill in the correct information based on the case.

Case 2
Female, 80. Fell in bathroom on hard floor, hurt spine and legs. High blood pressure, other vital parameters are fine.

Tasks:
You will be granted 1 main tasks. At any point, the decision support system might give you a notification.

When given a notification, look through the current information on the decision support page. Consider making a decision based on the data provided, or filling in important missing variables. If the latter, start with the most important variable.

1. On the **WEST** page, look through the vitals and check that they are ok. Fill in **Situation** information.

C

Consent form: Usability test

Please read and sign this form.

During this usability test I agree to participate in the testing of a digital prototype of an ambulance it-system. During the session my voice and hands will be recorded. I will be asked to find information or complete tasks and verbalize my thoughts while I do so. I will then be interviewed with follow-up questions and asked to complete a usability survey.

I understand and consent that the test leaders David Wallstén and Gregory Axton will use the recording for research purposes only and not spread it further. The data from the recording will be analyzed and anonymized by the test leaders for further use.

I understand that participation is voluntary, and I agree to immediately raise any concerns I might have.

Please sign below to indicate that you have read and understand the information on this form and that any questions you might have about the session have been answered.

Date:

Please print your name:

Please sign your name:

D

SUS results

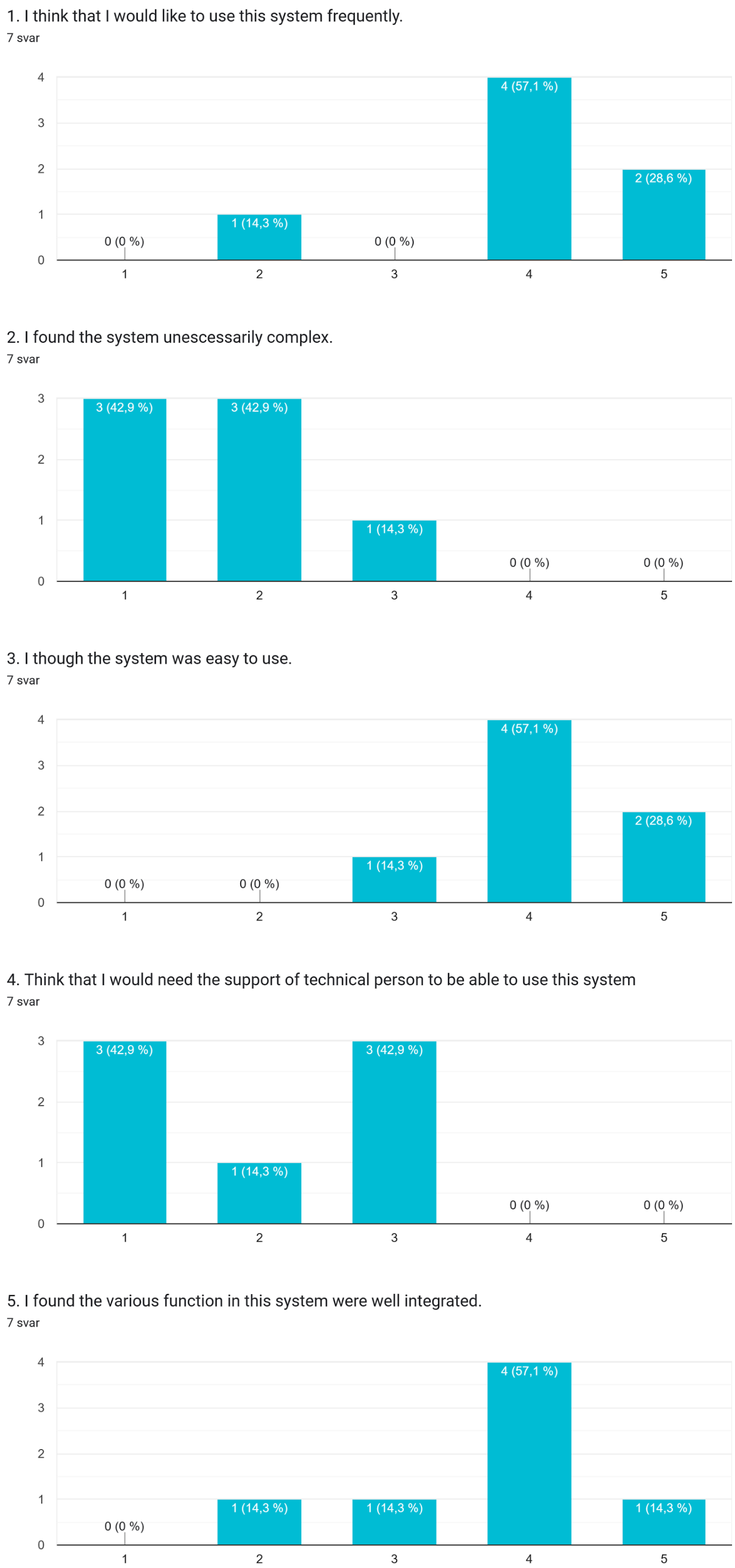


Figure D.1: SUS results 1-5

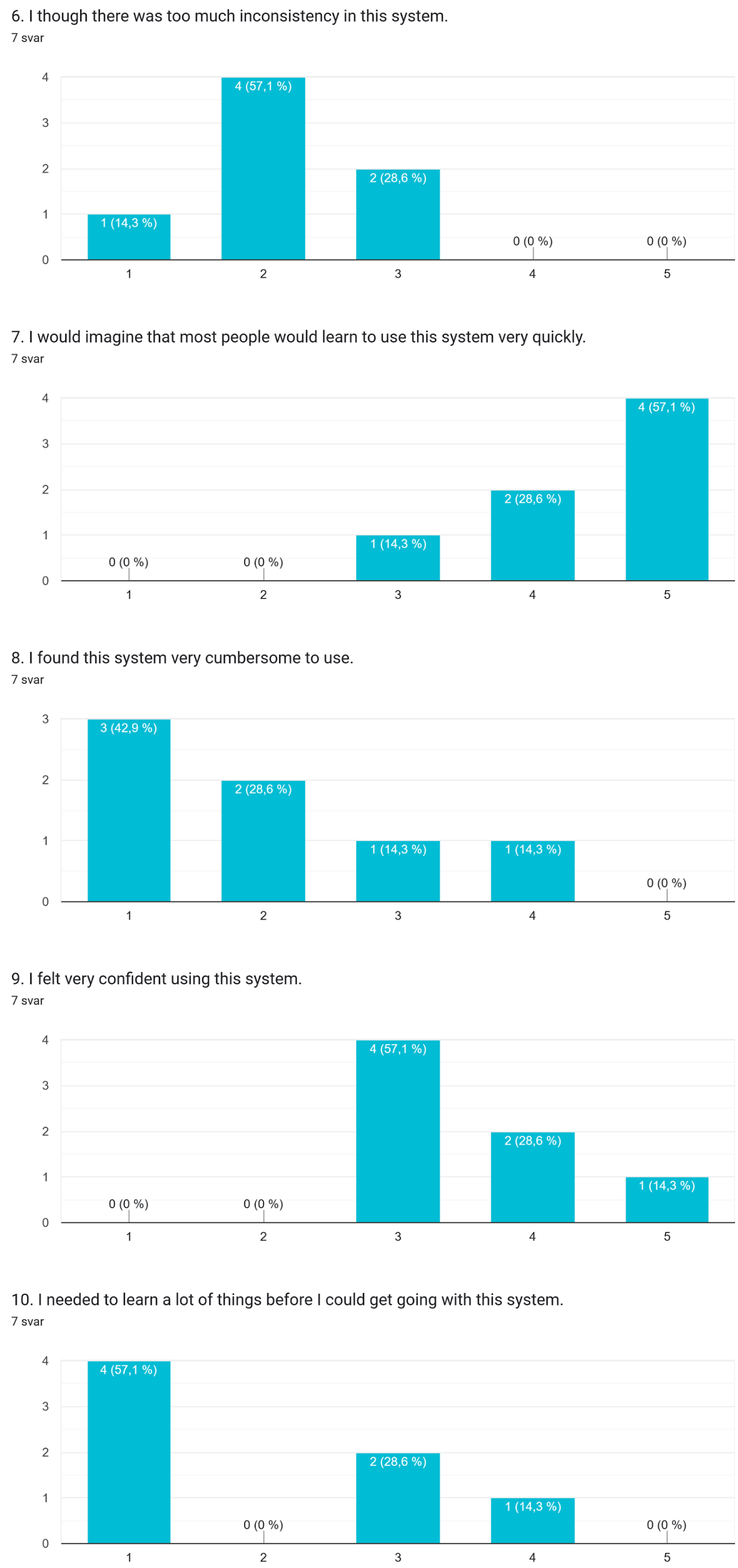


Figure D.2: SUS results 6-10