

Philip Hodne, Oskar K. Skåden

**NTNU**  
Norwegian University of  
Science and Technology  
Faculty of Architecture and Design  
Department of Design

Philip Hodne  
Oskar K. Skåden

# Automation Transparency For Maritime Autonomous Surface Ships Through Conversational User Interfaces

June 2022





Norwegian University of  
Science and Technology

# Automation Transparency For Maritime Autonomous Surface Ships Through Conversational User Interfaces

**Philip Hodne**

**Oskar K. Skåden**

Industrial Design

Submission date: June 2022

Supervisor: Ole Andreas Alsos

Co-supervisor: Thomas Porathe

Norwegian University of Science and Technology  
Department of Design



# Automation Transparency For Maritime Autonomous Surface Ships Through Conversational User Interfaces

By

Philip Hodne and Oskar K. Skåden

## Abstract

The increased autonomy in maritime shipping is expected to make the industry safer, more efficient, and reduce its environmental impact. Still, efficient interactions between autonomous agents and humans are fundamental to increasing trust and reducing the risk of collision. One challenge for Maritime Autonomous Surface Ships (MASS) is facilitating interactions through radio between ships and shore stations. Despite the importance of VHF for safe maritime traffic, current MASS prototypes are not able to understand nor participate in this communication. This thesis proposes using a conversational user interface (CI) with the autonomous agent in MASS as a replacement for radio in vessel-to-vessel and vessel-to-infrastructure communication. Based on maritime communication guidelines and textual descriptions of a MASS' surroundings, we demonstrate the use of the transformer-based deep learning model GPT-3 to enable human-like answers to prompts of status, decisions, and future intent. Usability tests with ship officers and Vessel Traffic Service operators (N=9) suggest that in peer-to-peer conversations, human operators are able to obtain a perceived sense of situational awareness from the MASS through the CI. A discussion of the opportunities and limitations of CI's in a maritime shipping environment saturated by autonomous systems concludes the thesis.

## Preface

This thesis was written in the Spring of 2022 as the concluding part of our master's degree programme in Industrial Design at The Norwegian University of Science and Technology (NTNU). We both have experience working within the maritime industry, making this a unique opportunity to utilise and build upon our acquired knowledge. The development of this thesis was done in collaboration with SFI Autoship, a Norwegian research-based innovation centre that will develop and manage technologies, systems, and operations for safe, sustainable, secure, and cost-effective autonomous sea transport operations. It is hosted by NTNU and involves six departments from three faculties and 22 industry partners. These include ship operators, shipping companies, suppliers of ship equipment and technology, ship classification and insurance, regulators, and researchers.

We want to thank our supervisors, Professor Ole Andreas Alsos and Professor Thomas Porathe from the Department of Design for ideas, insight, and inspiration along the way. Their supervision and thorough guidance have been invaluable throughout the semester. Finally, our gratitude extends to everyone who has helped with and participated in the project, especially Ketil Olaf Paulsen of Kongsberg Maritime. Their contribution was indispensable.

Trondheim, June 2022

Philip Hodne and Oskar K. Skåden

## Table of contents

1 Introduction	4
1.1 Human factors	5
1.2 Explainable AI	5
1.3 Automation transparency	6
2 Background	8
2.1 Stakeholders MASS must communicate with	8
2.2 The maritime radio communication challenge	9
2.3 Radio communication frameworks	11
2.4 Conversational user interfaces	13
2.5 Natural language processing	14
2.6 GPT-3	15
3 Aim of research	16
4 Method	16
4.1 Human-centered design	16
4.2 Understanding the context	17
4.3 Specifying the user requirements	19
4.4 Producing design solutions	19
4.5 Evaluating the designs	23
5 Results	27
5.1 Problem highlights and user needs	27
5.2 Input from the Subject Matter Experts	29
5.3 Testing the language model	30
5.4 Building the prototype	33
5.5 Usability test results	39
5.6 System Usability Scale Results	45
6 Discussion	46
6.1 Should you be able to negotiate through the CI?	48
6.2 Technical implementation	48
6.3 Limitations	50
7 Conclusion	52
References	53
Appendix	60

## Table of figures

Figure 1: XAI concept.	6
Figure 2: Automation transparency makes MASS' rationale available.	7
Figure 3: MASS with object awareness and future intention shown on ECDIS.	7
Figure 4: Animated signs on an autonomous urban ferry sharing status and intention.	7
Figure 5: Simplified model of the stakeholders MASS must communicate with at sea.	9
Figure 6: Proposed protocol for collaborative collision avoidance.	10
Figure 7: Estimated cost-benefit of automating missing maritime communication for MASS.	11
Figure 8: IALA VTS VHF guidelines on sentence structuring.	13
Figure 9: System overview of a CI.	14
Figure 10: The HCD process for our proposal.	17
Figure 11: From the ship bridge (left) and the desk of the VTS operator (right).	18
Figure 12: The three guiding constraints for our proposal.	19
Figure 13: From sketches to prototypes.	20
Figure 14: Proposed model for how information about the world flows into the CI.	21
Figure 15: Coding the web application.	22
Figure 16: The collision between Scot Carrier and Karin Høj.	24
Figure 17: Our adaption of the Scot Carrier and Karin Høj collision.	25
Figure 18: Our adaption of the Godafoss accident with Automat Express sailing northbound.	26
Figure 19: System overview of the CI communication.	33
Figure 20: The CI system's input and output.	34
Figure 21: CI system embedded in an ECDIS.	35
Figure 22: SMCP message markers with translations may be added for convenience.	35
Figure 23: The two systems and the urgent call failsafe switch.	36
Figure 24: Group broadcasts would enable multiple ships to coordinate.	37
Figure 25: Information architecture for the CI system web application.	37
Figure 26: The web application with the overlay chat active.	38
Figure 27: Identification of vessel, and, the microinteraction when the press-to-talk is active.	38

## Table of tables

Table 1: IMOs levels of autonomy for MASS	4
Table 2: SMCP message markers	12
Table 3: Subject Matter Experts interviewed	20
Table 4: Participants in the usability test	23
Table 5: Needs and goals of the stakeholders	28



## Abbreviations

AI: Artificial intelligence

AIS: Automated Identification System

CI: Conversational User Interface

COLREGS: International Regulations for Preventing Collisions at Sea

DSC: Digital Selective Calling

ECDIS: Electronic Chart Display Information System

HMI: Human-machine Interface

IALA: International Association of Maritime Aids to Navigation and Lighthouse Authorities

IMO: International Maritime Organization

ISO: International Organization of Standardization

MASS: Maritime Autonomous Surface Ships

MUNIN: Maritime Unmanned Navigation through Intelligence in Networks

ROC: Remote Operating Centre

SOLAS: International Convention for the Safety of Life at Sea

SMCP: Standard Marine Communication Phrases

V2I: Vessel-to-infrastructure

V2V: Vessel-to-vessel

VDES: VHF Data Exchange System

VHF: Very High Frequency

VTS: Vessel Traffic Services

XAI: Explainable AI

# 1 Introduction

Being able to automate workflows can increase productivity and lead to superior cost-efficiency. Maritime Autonomous Surface Ships (MASS) is one technology aiming to automate parts of the maritime shipping industry. Enabling the automation of the maritime industry has been an ongoing effort for several years (IMO, 2021). The International Maritime Organisation (IMO), with its central role in guiding the development, defines MASS as a “ship which, to a varying degree, can operate independent of human interaction”. The definition does not, however, define a specific type of ship. A key component of IMOs definition is that it encompasses varying degrees of autonomy. There are multiple proposed taxonomies for levels of autonomy in MASS. Still, IMOs is the most referenced in current research (Veitch & Alsos, 2022). As seen in table 1, IMO has defined four levels of autonomy for ships, where each level has a lower degree of human influence (IMO, 2021). For the remainder of this thesis, we will focus mainly on the latter two categories, defining MASS as a vessel without seafarers on board. This represents the biggest change from today’s system.

Table 1: IMOs levels of autonomy for MASS

<b>Degree 1</b>	Ship with automated processes and decision support
<b>Degree 2</b>	Remotely controlled ship with seafarers onboard
<b>Degree 3</b>	Remotely controlled ship without seafarers onboard
<b>Degree 4</b>	Fully autonomous ship

An important factor in automating the shipping industry is artificial intelligence (AI). AI may be defined as computational algorithms that aim to resemble human cognition and intelligence. The algorithms are programmed to recognise patterns and carry out predictions, decisions, and operations with speed and precision (Nilsson, 1998, pp. 1-8). For MASS, the AI is to replace and improve upon the intelligence of today’s human crew. As AI and automation technology matures, and vessels are moving up the degrees of autonomy, some ship officers are expected to move from the ship bridge to shore where they will monitor multiple autonomous ships simultaneously from a remote operating centre (ROC). Despite no formal definition, ROCs, also known as Shore Control Centres, describes control rooms where humans can command, control, and monitor autonomous ships through electronic equipment. The hypothesis is that one captain will be able to control 6 ships simultaneously (MacKinnon *et al.*, 2015; Dybvik, Veitch & Steinert, 2020). When the captain is moved to the ROC, there are different proposals for the engagement of this human operator (Veitch & Alsos, 2022), however, a primary goal is for them to be a passive entity. That is, the ships will navigate autonomously, and the role of the ROC operator is characterised by supervision and pre-emptive actions. Only in exceptional cases is the ROC operator expected to intervene if alerted by the autonomous system (Porathe, Fjortoft & Bratbergsengen, 2020). Fully autonomous, independent MASS is arguably a common goal for autonomous sea transport operations. Nevertheless, some, such as Relling *et al.* (2018), argue that this is neither achievable nor desirable and should therefore not be the objective. The push to omit humans and leave the navigation to a set of AI algorithms is driven by an expectation of safer, more competitive, and more environmentally friendly shipping alternatives (Rødseth, 2018; Kim & Schröder-Hinrichs, 2021). Especially in the short-sea shipping segment is the concept of autonomous vessels attractive. Suffering from pollution, congestion, and population growth in urban areas, the demand from multiple governments to move transportation from roads to the sea is growing. The increased profitability and

the expected reduction in greenhouse gas emissions from using MASS are expected to help accelerate this transition (Rødseth & Burmeister, 2015a; Tvette, 2015).

Still, the deployment of MASS is in its infancy. Despite research funding and effort, the progress remains slow, with commercial uptake being largely non-existent (Chae, Kim & Kim, 2020). The wave of research started with the MUNIN project running from 2012 to 2015 (Burmeister & Rødseth, 2013), followed by DNVs ReVolt project presented in 2014 (Tvette, 2015). More recently, the KASS project will research core technologies for autonomous ships towards 2025 (KASS, 2020), while the three-year AEGIS consortium (2020) will deliver research on connected, autonomous waterborne logistics systems. Yara Birkeland unveiled in 2017 as the world’s first electric autonomous container ship made its first voyage in April 2022. The ship will be crewed through a two-year testing period before moving the captain to an onshore ROC (Yara, 2022). Indeed, there are challenges inhibiting the mass utilisation and deployment of MASS. These include, amongst others, a lack of regulatory frameworks (Faivre *et al.*, 2020; Ringbom, Røsæg & Solvang, 2021), inadequate monitoring, navigation, and collision avoidance systems (Zhang *et al.*, 2021; Kim & Schröder-Hinrichs, 2021), as well as concerns regarding cybersecurity (Chang *et al.*, 2020; Bolbot *et al.*, 2020; Bolbot *et al.*, 2021) and human factors (Burmeister *et al.*, 2014; Rødseth & Burmeister, 2015b; Chang *et al.*, 2020). We refer to Veitch & Alsos (2022) for further elaborations on the research gaps, regulatory obstacles, and technical shortcomings of MASS development.

## 1.1 Human factors

In the previous section, we identified human factors as a barrier to the real-world implementation of MASS. Human factors is a scientific discipline concerned with understanding the interaction between people and technology and making that interaction more efficient, safer, and more pleasant. It examines the latent causes of errors and factors that can improve these interactions (Stanton *et al.*, 2005, pp. 1-3). For MASS, the safe interaction between the autonomous system and humans is prominent (Porathe, 2019a; Liu *et al.*, 2022). Chang *et al.* (2020) conclude that it is one of the most significant contributors to the overall risk of MASS operations. As an increasing number of MASS are deployed in safety-critical environments at sea, behaving based on a set of predictions, the ability for humans to understand the autonomous vessels’ status, decisions, and intent is critical for establishing trust and reducing the risk of collisions (Porathe, 2019a; Porathe, 2021a). The effects of collisions are multifactorial and include the potential loss of human lives, environmental damage, and costly operational disturbances to the shipping industry (Karahalios, 2014). This need for assurance illuminates a broader challenge within the field of AI. Explainable AI (XAI) is a research field aiming to achieve transparency between autonomous systems and humans. In the next section, we will outline the concept of XAI and how it may be vital to solving the aforementioned human factors risk.

## 1.2 Explainable AI

Machine learning and other AI techniques have seen success in recent years, fuelling their application in our everyday lives. Using prediction models based on collected training data, AI algorithms can predict what output should be given without explicitly being programmed to do so. While today’s best performing AI models have high predictive accuracy, this comes at a cost; their explainability. The more complicated the decision, the less clarity they can provide on how they arrived at their result. It is difficult for humans to understand which of the several inputs fed to the model at the beginning are driving the result. This “opaque-box” nature of AI techniques has given rise to the field of XAI. XAI

aims to create a suite of techniques that enable the developers and end-users to understand, trust, and manage AI models. DARPA (Defense Advanced Research Projects Agency) defines XAI as “AI systems that can explain their rationale to a human user, characterize their strengths and weaknesses, and convey an understanding of how they will behave in the future.” (Gunning & Aha, 2019). This push is in part driven by a perceived growing importance of AI ethics, with the frequent occurrence of calls for privacy, non-maleficence, justice, and fairness from both public entities and private enterprises (Jobin, Ienca & Vayena, 2019). The focus is twofold; the design of the models themselves and the human-machine interfaces (HMI) mediating the process and output. HMIs may be defined as screens or other technologies that connect a human to a machine and allows for interaction. As seen in Figure 1, an explainable AI model will receive input and deliver the results to the end-user through the HMI. Streamlining this interaction is thus a central part of XAI development (Gunning, 2016).

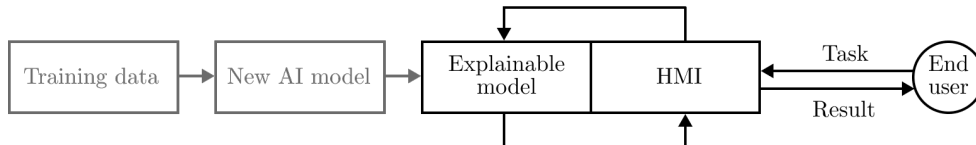


Figure 1: XAI concept.

Meske *et al.* (2020) propose a framework for how the HMI of an autonomous system should communicate. Explanations should be easy to receive, understandable for the individual user, context-specific rather than generic, and personalised to the user, all delivered with a high degree of confidence and information value. An autonomous agent built upon the principles of XAI should allow the user to request information about how a particular outcome was reached, and which aspects of input data and other parameters of the system influenced its decision or prediction. The user is then able to assess the appropriateness of the outcome, which can help strengthen trust and confidence in the autonomous system (Ras, van Gerven & Haselager, 2018).

### 1.3 Automation transparency

A focus on XAI may help solve human factors challenges within MASS. Within the field of autonomous machines and robotics, the term *automation transparency* may be used to describe the XAI HMI’s ability to be understandable. In his model on automation transparency, Lyons (2013) defines the term as a property enabling humans to “correctly perceive the ability, intent, and situational constraints of the automation or autonomous system”. An end-user is inherently blind to the system’s reasoning unless deliberate design decisions are taken. By focusing on automation transparency, one grants an end-user access to a closed system’s rationale. Lyons argues that designing an interface that increases the degree of automation transparency can be done through various levels. This includes the informational level with the use of light, symbols, and text, the communicative level with the use of sounds and voice, and the physical level through movement and gestures.

As outlined previously, MASS will sail and navigate by themselves without human intervention. A longstanding challenge within MASS research is the interaction between autonomous and crewed vessels. One aspect is that today’s MASS prototypes only communicate through movement. Data beyond movement - their perception of the world and decisions thereafter - is only accessible to the operators in the ROC (Dybvik, Veitch & Steinert, 2020; Alsos *et al.*, 2022). For the human stakeholders outside the ROC that are not part of the operation, yet affected by it, the lack of automation transparency makes collaboration and interaction with the autonomous vessels arduous (Porathe, 2015; Porathe,

2019a). Thus, much effort has been placed into making the status, decisions, and intent of the MASS available and transparent for other ships and shore stations (see Figure 2) (Porathe, 2019a; Porathe, 2019b; Porathe & Rødseth, 2019; Porathe, Fjortoft & Bratbergsengen, 2020; Relling *et al.*, 2021; Alsos *et al.*, 2022).

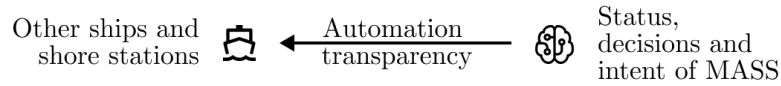


Figure 2: Automation transparency makes MASS' rationale available.

Previous research has proposed, amongst others, increasing automation transparency through the informational level by adapting navigation equipment to share future intent and status, known as Route Exchange (Porathe, 2019b) (seen in Figure 3), and adding custom markings and messages for MASS on the Automatic Identification Systems (AIS) and Electronic Chart Display and Information Systems (ECDIS) of ships (Alsos *et al.*, 2022). As seen in Figure 4, embedding lights on the superstructure of the ship, enabling the broadcast of status and intent has also been proposed (Porathe, 2021b).

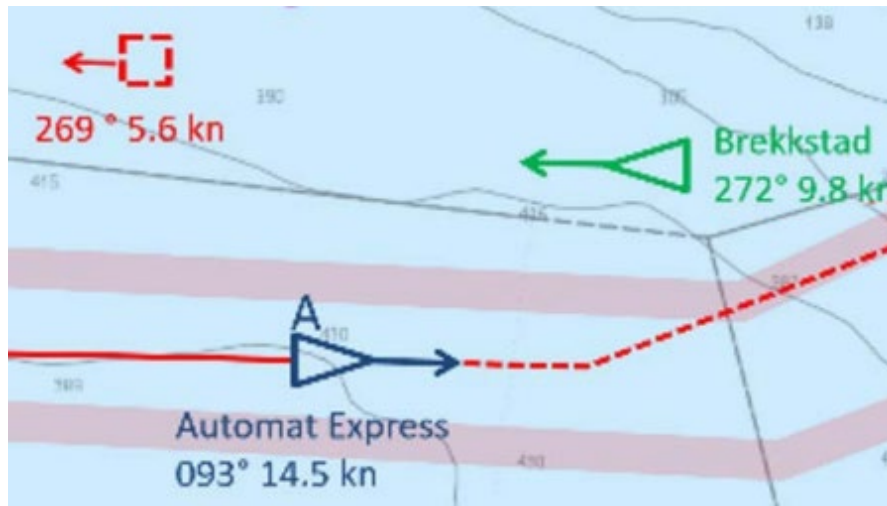


Figure 3: MASS with object awareness and future intention shown on ECDIS.

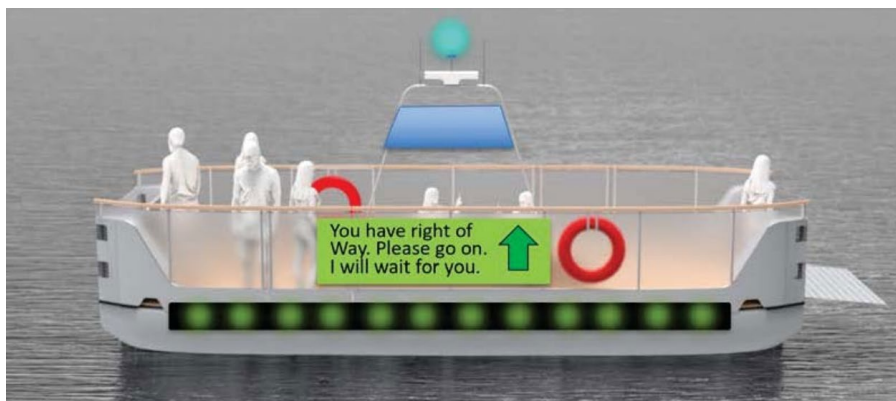


Figure 4: Animated signs on an autonomous urban ferry sharing status and intention.

The most important factor in making other vessels trust MASS is through the physical level, that is, its movement (Alsos *et al.*, 2022). The main body of research has been on technology making MASS sail in compliance with rules and mariner’s expectations (Wróbel, Gil & Montewka, 2020; Zhang *et al.* 2021). Still, researching the prospect of automation transparency between autonomous and crewed vessels, Porathe (2019a) explains how the rules of the sea, such as the *Convention on the International Regulations for Preventing Collisions at Sea* (COLREGs), are ambiguous, requiring ship officers to evaluate several ship interactions on a case-to-case basis. Indeed, empirical evidence from Rutledal, Relling & Resnes (2020) showed how a significant number of evasive manoeuvres by ferries outside Ålesund, Norway defied the COLREGs to avoid possible collision situations. These manoeuvres are challenging to resolve without clarification beyond movement. Some situations are even resolved by «common sense manoeuvres», such as giving way to larger vessels despite the right to proceed (Akdağ, Solnør & Johansen, 2022). Still, in these cases, where rules are ambiguous and collisions are avoided based on the circumstances, humans commonly coordinate using radio. An ongoing research challenge for MASS is that current maritime radio communication is only tailored to humans. Today’s MASS prototypes are unable to participate in this communication. They are neither able to understand what is said nor able to express themselves – ultimately making them a liability at sea (Alsos *et al.*, 2022). We call this the “maritime radio communication challenge”. It is to this challenge we now turn.

## 2 Background

This section provides a background on the maritime radio communication challenge for MASS and the building blocks of our proposed conversational user interface. Our previous work in Alsos *et al.* (2022) provides an elaboration on this challenge and forms the basis of this section. This background section may be split into two parts. In the first part, we define the human stakeholders MASS must be able to communicate with when at sea. We then explain how MASS’ inability to understand and participate in conversations through radio is the largest bottleneck for its automation transparency, and a considerable risk factor in the safe interaction with crewed ships. In the second part, we outline the standards and regulations that provide the framework for radio communication in the maritime domain, the history and concept of conversational user interfaces, and, finally, an outline of natural language processing and the machine learning model GPT-3.

### 2.1 Stakeholders MASS must communicate with

In *1.3 Automation transparency*, we identified two stakeholder groups: the people running MASS operations from the ROC, and the ones affected by the operation. In this section, we will add granularity and context to the latter grouping by outlining who these stakeholders are. The stakeholders of MASS and their importance will to some extent vary based on the ship's purpose and the environment in which it sails. As mentioned, the largest potential for MASS operations is in the short-sea shipping segment, thus, we establish our initial constraint for our proposal by focusing on the stakeholders MASS must interact with in coastal waters.

While MASS would primarily interact with SOLAS ships while sailing in international waters, the stakeholders closer to shore in coastal waters and inland waterways are more numerous. In coastal waters, you may divide the stakeholder affected by MASS into 4 groups: SOLAS ships, non-SOLAS ships, Vessel Traffic Services (VTS), and port authorities. For the remainder of this paper, we refer to the first two as vessel to vessel (V2V) communication while the two latter as vessel to infrastructure

(V2I) communication (Figure 5). SOLAS (Safety of Life at Sea) is a convention that specifies minimum standards for the construction, equipment, and operation of ships under its jurisdiction (IMO, 1974). These are the vessels that transport, or support the transport, of people and goods with trained crew and standardised equipment such as ECDIS, radar, and AIS. We use the term SOLAS ships to refer to those regulated under this convention, which constitute 99.89% of the world’s merchant fleet (IMO, 2022). Non-SOLAS ships refer to ships not regulated by SOLAS. Here we find smaller leisure boats like sailboats and motorboats. They usually do not have the same sophisticated navigation and communication tools as SOLAS vessels. Often, they are sailed by less experienced mariners with equipment limited to a VHF radio and perhaps rudimentary AIS. The key differentiator is the requirement for communication equipment between SOLAS and non-SOLAS, where MASS must assume limitations for non-SOLAS. We refer to Alsos *et al.* (2022) for further analysis of vessel-to-vessel communication in the context of MASS. Furthermore, MASS must be able to communicate with the infrastructure at shore, most prominently VTS and port authorities, whose responsibility it is to ensure safe navigation in coastal areas and ports by means of monitoring, informing, and organising the maritime traffic (IMO, 1997). VTS and port authorities work similarly, providing information services, navigation assistance, and traffic organisation services to ships in its vicinity over VHF. The general difference is that the port authority manages one or more ports while VTS manages a larger coastal area.

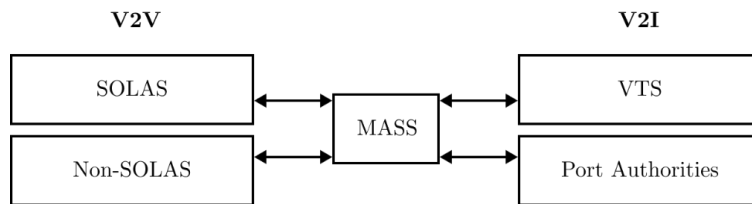


Figure 5: Simplified model of the stakeholders MASS must communicate with at sea.

Still, there may be more stakeholders that expect the MASS to communicate with them, such as crane operators in ports, bridge operators, passengers, and emergency services. While these stakeholders must be considered in the development of MASS, they are out of scope in this thesis as this is a simplified model of the stakeholders an autonomous vessel would communicate with over VHF most often. We argue that this model, if not exhaustive, will be comprehensive enough for further analysis.

## 2.2 The maritime radio communication challenge

Today, the safe interaction between traditional ships and shore stations is primarily facilitated by ship crew observing other ships’ movement visually and through AIS, radar, and other electronic equipment. The interaction is supported by a collective agreement that every ship follows the COLREGs, with monitoring and support by the VTS. Today’s MASS prototypes perceive other ships’ movement and navigate accordingly. The information about its surroundings and associated predictions is collected from sensors and aggregated in a process referred to as *sensor fusion*. This data is then combined to give the MASS a representation of its world, upon which it makes its predictions. Several inputs – AIS, radar, and camera vision – are fused together to create a prediction for what an object may be and how a MASS should carry out associated actions (Brekke *et al.*, 2019; Brekke *et al.*, 2022). In the future, it is expected that MASS will collaboratively plan and exchange their future routes in a shared network – machine to machine. It is expected that these collision avoidance algorithms will exchange and plan routes with participating vessels and, if needed, employ reactive collision avoidance to vessels that do

not act according to the plan; temporarily deviating from the intended route before returning (Akdağ, Solnør, & Johansen, 2022). An illustration of this system can be seen in Figure 6. In these situations, where a MASS must employ reactive collision avoidance, they would act in solitude without explicit collaboration with the non-participating vessel. In principle, they do not express themselves in any other way than by movement and act without announcements (Alsos *et al.*, 2022).

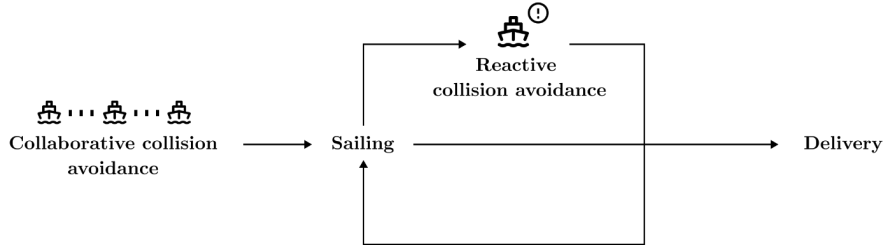


Figure 6: Proposed protocol for collaborative collision avoidance.

For humans on traditional ships that want to avoid collisions, these situations are commonly resolved through verbal communication. This way, humans can inquire about the other vessel’s intent and resolve the situation through dialogue. Radio is the standard tool used to solve confusion and prevent dangerous situations at sea. The radio technology used at sea is in large parts VHF radio, but MF and HF radio technology is also used to a lesser extent on international waters. These work similarly, however, in this thesis, we will focus on VHF given its more frequent use in coastal waters.

### 2.2.1 The use of VHF radio

VHF (Very High Frequency) radios consist of transceivers and receivers that are designed to send and receive radio messages on a designated band of frequencies. Regulation stipulates that all VTS stations and SOLAS ships must have a VHF radio on board, and it must always be on to listen to distress signals. For Non-SOLAS ships, it is not required to utilise a VHF radio, though most leisure boats of a larger size in coastal waters do. All VHF receivers are by default listening to one frequency called Channel 16. Channel 16 is intended for distress signals and routine calls. When speaking, English is to be used as your voice is broadcasted to all surrounding VHF receivers within a 50-70 nautical mile radius. DSC (Digital Selective Calling) technology enables you to call a specific ship using their Maritime Mobile Service Identity (MMSI) number and switch to a separate radio channel, similar to a telephone call. When entering an area monitored by a VTS or port authority, all communication should in most cases be switched to that shore station’s designated frequency, such as Channel 8. A rule of thumb is that VHF channels be kept open, and no unnecessary communication is conducted. As everyone is sharing the same channel, all conversations are heard by surrounding ships and shore stations. An important property of radio communication is the ability to quickly contact a nearby ship to clarify situations and solve confusion. This may be done in a separate DSC call, or over the shared channel (Du Port, 2021). To avoid misuse and confusion, maritime regulating bodies have guidelines and regulations for how conversations over VHF should be conducted. We will outline these in the *2.3 Maritime radio communication frameworks*.

### 2.2.2 The issue with VHF for MASS and their ROC operators

Despite guidelines not to use VHF radio without just cause, it is still necessary by regulation and expectation that ships are able to communicate with it. Especially in high traffic areas, such as coastal waters, is the use of radio important to quickly clarify intent and avoid accidents. Without any technology to perceive nor express themselves through radio, introducing MASS into today’s maritime



environment is prone to exacerbate safety at sea and increase collision risk (Alsos *et al.*, 2022). Until now, the solution was to relay the VHF communication to the ROC to be monitored and operated by humans. However, with ROC operators responsible for multiple ships, radio communication of up to 6 ships would have to be monitored and answered simultaneously. The worry is that this may exacerbate ROC operators’ concentration, strain their cognitive load, and, in the worst case, lead to collisions (Porathe, 2021A). As Dybvik, Veitch & Steinert (2020) summarises in their study of challenges for Shore Control Centers (SCC):

Effective communication was highlighted by several interviewees as an essential component of the SCC design. [...] One interviewee pointed out that modern regulations stipulate continuously being tuned in to VHF-radio onboard each ship, which would make communication chaotic when operating several vessels simultaneously. This was also mentioned by several other interviewees; having more than one communication channel would be an issue. They questioned how many vessels one can communicate with.

Indeed, in a future with the proposed collaborative collision avoidance systems the use of VHF radio to avoid collisions would be redundant as all routes are planned by the participating vessels. Still, this assumes the inclusion of all vessels at sea throughout their journeys, which is not a reality. The collaboration between humans and autonomy is still an important aspect of the future of MASS as mariners on non-SOLAS and SOLAS ships that are not participating will continue to be part of the maritime ecosystem (Akdağ, Solnør & Johansen, 2022). An analysis of the communication methods MASS is not able to participate in was conducted in our previous work (Alsos *et al.*, 2022). Here, we identified radio as one of the most important communication methods to make MASS adapt to for safe and efficient operations (see Figure 7).

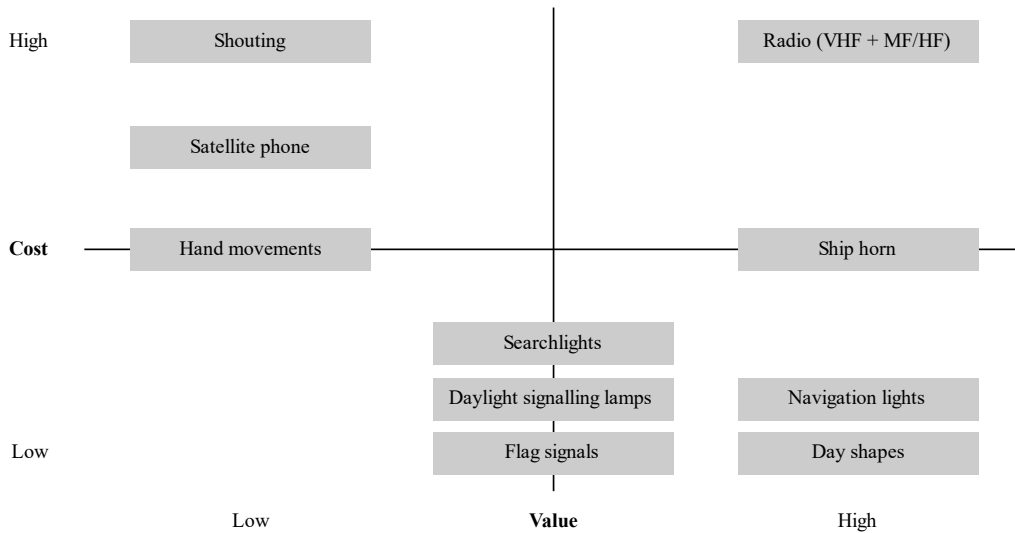


Figure 7: Estimated cost-benefit of automating missing maritime communication for MASS.

### 2.3 Radio communication frameworks

In the last section we explained what tasks are done over VHF, and how deploying MASS’ without the ability to participate in these tasks would be hazardous. We also outlined how the adoption of regulation to avoid communication problems over radio has been put into place. In this section, we summarise the existing regulations and guidelines for radio communication that are widely deployed in the maritime environment. These frameworks provide guidelines and influence the direction of our proposal.

### 2.3.1 Standard Marine Communication Phrases (SMCP)

IMO acknowledges the importance of clear and unambiguous communication at sea and has taken steps to improve maritime communication by standardising it through resolutions. One of these resolutions is the Standard Marine Communication Phrases (SMCP) which was adopted in 2001. The International Convention on Standards of Training, Certification and Watchkeeping for Seafarers (STCW), by which 99.03% of the world’s merchant fleet is regulated, requires that all officers in charge of a navigational watch are able to use and understand the content of the resolution (IMO, 2001). SMCP broadly divides communication phrases into external communication phrases and on-board communication phrases. According to IMO (2001) these phrases are constructed using standardised structures in simplified English, with the intention to reduce grammatical, lexical, and idiomatic varieties. One could say SMCP simplifies the language without reducing clarity, thus potentially reducing the barriers posed by language. IMO summarises the features in SMCP’s communicative phrases as:

- Avoiding synonyms
- Avoiding contracted forms
- Providing fully worded answers to “yes/no”-questions and basic alternative
- Answers to sentence questions
- Providing one phrase for one event, and
- Structuring the corresponding phrases according to the principle: identical invariable plus variable.

IMO also defines what they call message markers which can be used to specify the nature of a message. IMO divides messages into the following 8 categories:

Table 2: SMCP message markers

Instruction	Advice	Warning	Information	Question	Answer	Request	Intention
-------------	--------	---------	-------------	----------	--------	---------	-----------

SMCP also provides guidance for VTS specifically on the basis that “The efficiency of a VTS will depend on the reliability and continuity of communications and on the ability to provide good and unambiguous information.”. It must be noted that while many officers are required to understand SMCP, they are not obligated to use it.

### 2.3.2 IALA Guidelines

International Association of Maritime Aids to Navigation and Lighthouse Authorities (IALA), which provides guidelines to shore infrastructure such as VTS and port authorities, has also recognised the need for communication guidelines. In their *VTS VHF Voice Communication Guidelines*, IALA remarks that sentences should be as short and concise as possible to prevent ambiguity (IALA, 2022). In this manner, they adopt the same message markers as the SMCP. A flowchart is illustrated in Figure 8, showing the guidelines for how a conversation through VHF should be conducted by VTS operators (IALA, 2017).

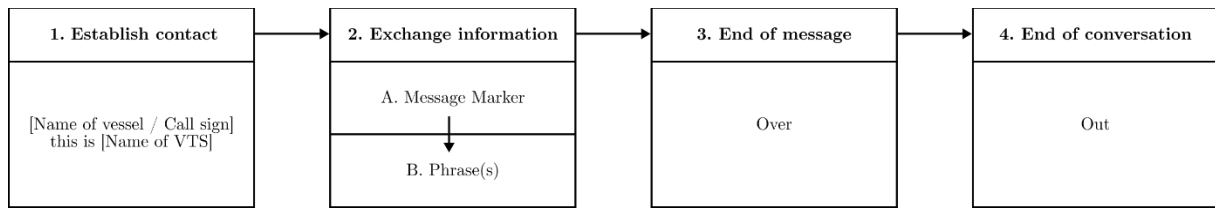


Figure 8: IALA VTS VHF guidelines on sentence structuring.

IALA advises VTS operators to always communicate in accordance with communication procedures to ensure professionalism and clarity, following the guidelines for sentence structures and vocabulary.

### 2.3.3 Critique of frameworks and use of VHF

While standardisation and guidelines may provide a good foundation for efficient communication in the maritime domain, their efficacy largely relies on whether people follow and understand them. A study by Nilsson (2013), based on 168 hours of recorded VHF communication between a VTS and ships within its Ship Reporting Area, found that SMCP was rarely used and that misunderstandings seldom occurred. These results are supported by Boström (2020), that found in his study of 40 days of ship-to-ship communication a clear gap between stipulated communication protocol and actual language use. The conclusion by Nilsson (2013) from his study was that communication in the maritime domain is “too complex to be adequately pre-scripted” (p. 58). Even though misunderstandings were rare in Nilsson’s study, the use of VHF itself has been identified as a root cause of accidents at sea (Bocanegra-Valle, 2011; Porathe, Eklund & Göransson, 2014; IALA, 2017). Communication problems between different actors in particular lead to a wrong assessment of traffic situations and subsequently result in accidents. These communication problems and the misinterpreting of information are major causes of marine casualties, accounting for about one-third of all human-induced accidents (Ziarati *et al.*, 2011). Unwanted background noise and the overall acoustic conditions on board ships, as well as the quality of the audio input itself, can make it further challenging to communicate effectively (John & Reimann, 2020). A proposal for text-based peer-to-peer messaging in non-time-critical situations has been made by Porathe, Eklund & Göransson (2014). In their study, they compared providing navigational instructions over VHF and text in a shipping simulator. The participants, four watch officers, and five 4<sup>th</sup> year cadet students, considered the text-based messaging system less stressful than voice calls since it didn’t require the participants to respond immediately. Based on these results they argue text-based communication has the advantage that people don’t need to remember instructions since messages are saved and can be read as many times as necessary. Moreover, they argue it reduces the risk of instructions being misinterpreted and of the receivers making unwarranted assumptions about its meaning. Porathe, Eklund and Göransson conclude that text-based messages can be a viable complement to traditional voice-based communication in situations that are not time-critical. Referencing this study, the research project ARTUS proposed the automatic transcription of VHF radio messages using AI-driven speech-to-text technology. Still, they theorised transcriptions using speech-to-text technology would be difficult to deploy given the quality of VHF radio and the noisy environment it is received (John & Reimann, 2020).

## 2.4 Conversational user interfaces

Conversational user interfaces (CIs), known also by terms such as spoken dialogue systems or conversational systems, are a type of HMI. They allow humans to interact with computers using natural spoken language, through speech or text. McTear (2002) defines the CI’s “as computer systems that use

spoken language to interact with users to accomplish a task”. The origins of CIs can be traced back to AI research in the 1950s (Allen, 2003) and, notably, the presentation of ELIZA in 1966 (Weizenbaum, 1966). ELIZA was one of the earliest examples of a computer program that allowed humans to talk to a computer program. The CI simulated conversations with a mock psychotherapist through a terminal. With developments in technology and increasing user acceptance and adoption in recent years, a surge in CI use can be observed; with systems such as voice assistants and chatbots being increasingly embedded in electronic equipment (McTear, Callejas, & Griol, 2016, p. 16). We describe the technologies empowering these systems in the next section, *2.5 Natural language processing*.

A system overview of CIs may be made based on a division of 6 functionalities that build the experience of conversing with the computer (Figure 9) (McTear, 2002):

- (1) Speech recognition: The user’s input speech is converted into text, if not sent that way already.
- (2) Language understanding: The string of words is analysed by the CI with the aim of understanding the complete meaning of the text.
- (3) Dialogue Management: The system and the user manage and coordinate their interaction, including ending the input phase and continuing the process.
- (4) Communication with external system: If needed, the system communicates with a database or other computer application to complete a task and craft an answer.
- (5) Response generation: The response is generated and sent back as text.
- (6) Speech output: The system’s text is converted into speech.

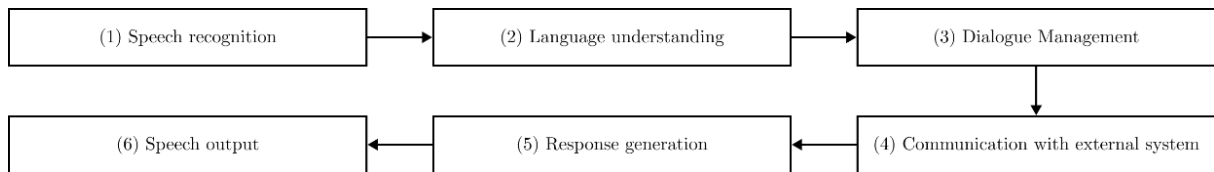


Figure 9: System overview of a CI.

McTear, Callejas, & Griol (2016, p. 11) comment that the chief value proposition of CIs is that they are a natural way for humans to interact with a system. Still, researchers claim the successful use of CIs is reliant on factors beyond just speech. Cassell’s (2001) remark – that “We need to locate intelligence, and this need poses problems for the invisible computer. The best example of located intelligence, of course, is the body.” – points to the unfulfilling user experience of interacting with a computer purely through text or speech without the embodied behaviours of humans. She stresses the importance of intelligent user interfaces that provides human representation in multiple modalities. She does not claim that all CIs must be represented as humanoids but argues that the CI systems should rely on the same interactional rules that humans do to organise and regulate the conversation. Cassell argues that pauses in speech, affirmative cues, and other conversational and gesture-based interactions help the user of a CI contextualise the information provided by the system.

## 2.5 Natural language processing

This section gives an overview of the field of Natural language processing (NLP) and introduces transformer-based deep learning models on which we base our proposal. NLP is a subfield of linguistics, computer science, and AI that is concerned with the ability of computers to understand and generate

human language, including the ability to hold conversations (Allen, 2003). The development in processing power and availability of vast amounts of data have contributed to the increased use of artificial neural networks (ANN) and deep learning. ANN are computing systems loosely based on the neural architecture of the brain with connected nodes, called artificial neurons, that are aggregated into layers. Deep learning is a subset of ANN referring to models that have a substantial number of these layers that process data. When training the ANN model on data, the parameters, or the coefficients of the artificial neurons in the layers between the input and output, are given different numerical weights by the model to give greater or lesser prominence to some aspects of the data in the overall calculation. The weighting of these parameters, and the amount of them, give the ANN its rationale for predictions. (LeCun, Bengio & Hinton, 2015). It is with these deep learning models that NLP has seen recent success, including development in text- and speech recognition, text analysis, and text generation (McTear, Callejas & Griol, 2016, p. 16). A principal difficulty in developing the field of NLP has been the pervasive ambiguity found at all levels of language; from lexical to semantic. For example, the word “duck” can be a noun [the animal] and a verb [to avoid something thrown]. Indeed, the ability to assume an implied context from previous statements is key to understanding humans’ natural language (Allen, 2003). Introduced in 2017, transformers are deep learning models that contributed to the advancement of NLP. The transformer models differ from other models in their improved ability to extract the meaning of a word based on its context in a sentence (Vaswani *et al.*, 2017). The introduction of transformer-based deep learning models led to the development of the BERT language model currently used in Google’s search engine (Devlin *et al.*, 2019), Microsoft’s Turing-NLG (Rosset, 2020), and the language model GPT-3. The next section outlines further the latter model and concludes the background section.

## 2.6 GPT-3

This section presents the NLP language model we used to build our prototype. Natural Generative Pre-Trained Transformer 3 (GPT-3) is a transformer-based deep learning model developed by the research laboratory Open AI in 2020. The third generation of this language model, it is proficient in processing, analysing, and generating text. GPT-3 has been pre-trained on text from the open internet. The data includes English-language Wikipedia, two datasets containing books, and 45 terabytes of plaintext from Common Crawl, giving the language model a total of 175 billion parameters (Brown *et al.*, 2020), 116 times larger than its predecessor GPT-2 with 1.5 billion (Solaiman *et al.*, 2019). The goal of GPT-3 is to intuit what task you are trying to perform based on a prompt and generate a plausible completion with few or no examples. This is referred to as "few-shot learning". Fine-tuning of the model is also possible, where fewer resources on smaller datasets can be used to optimize its performance on specific tasks. GPT-3 can be applied to solve tasks that involve processing language, including content generation, summarisation, semantic search, topic tagging, and sentiment analysis. (Brown *et al.*, 2020). Since its release, multiple transformer-based deep learning algorithms based on larger data sets and novel techniques have demonstrated superior capabilities<sup>1</sup>, notably Megatron-Turing NLG with 530 billion parameters (Smith *et al.*, 2022) and Switch Transformer with 1.6 trillion parameters (Fedus, Zoph & Shazeer, 2022). GPT-3 is available through an API, released in late 2021. APIs (Application Programming Interface) are interfaces that enable a computer to connect to a database and utilise its data. The source code of GPT-3 is not available, and the use of the model is priced. All applications using the GPT-3 API must be reviewed by Open AI before being released to the public (Open AI, 2022).

---

<sup>1</sup> As measured in parameters. One could argue the number of parameters is irrelevant if the datasets it is trained on is inferior.

### 3 Aim of research

In this thesis, we focus on adapting maritime communication to a future more saturated with autonomous vessels to ensure and improve safety at sea. This thesis proposes the use of a conversational user interface that will help increase the automation transparency of a MASS when operating with humans on nearby ships and shore stations. The conversational user interface is to work as a proxy to the mind of the autonomous agent, enabling mariners to access the MASS' state, decisions, and intent using natural language. The chief value lies in its aid in reactive collision avoidance situations and communication with sea traffic management infrastructure, such as VTS and port authorities. By automating this communication, you also help alleviate ROC operators' workload and help fulfil their role as a backup in the operation of MASS.

The research question the present thesis tries to answer with its proposal is: "How may we solve the maritime radio communication challenge, meeting the needs and goals of both human and MASS stakeholders?". To prepare the most feasible solution, we employed the use of qualitative research methods and processes from the field of human-centered design. This is outlined further in the next section.

### 4 Method

This chapter presents the framework we used to build and usability test our conversational user interface prototype. We outline human-centered design (HCD), the ISO standard model for the HCD process, and our procedures for data procurement, product development, and usability testing.

#### 4.1 Human-centered design

Human-centered design (HCD), often synonymous with the term user-centered design, is an approach to problem-solving focused on designing solutions from the user's point of view. The goal is to create solutions that empathise with the user, ultimately improving the usability, accessibility, and user experience of a system. With an emphasis on participation by the members of the system, the goals and needs of the users are brought into the innovation process to create solutions that are optimized to what the users want instead of forcing them to change behavior and expectations (Norman, 2013, p. 8). Norman refers to how humans represent these behaviors and expectations as the "mental model" for how a product or service works. ISO (the International Organization for Standardization) has provided in their ISO 9241-220:2019 *Ergonomics of human-system interaction* standards a process that they claim represent good practice for HCD within and across projects. In their model (Figure 10), an iterative approach to the design of solutions is proposed until the needs of the users are met. The model consists of four phases: (1) understanding the context, (2) specifying user needs, (3) producing designs, and (4) evaluating solutions. It is this general model we followed to align our proposal to the tenets of HCD. In the next sections, we will outline which tangible methods we used for each of these phases throughout the process of creating our proposal.

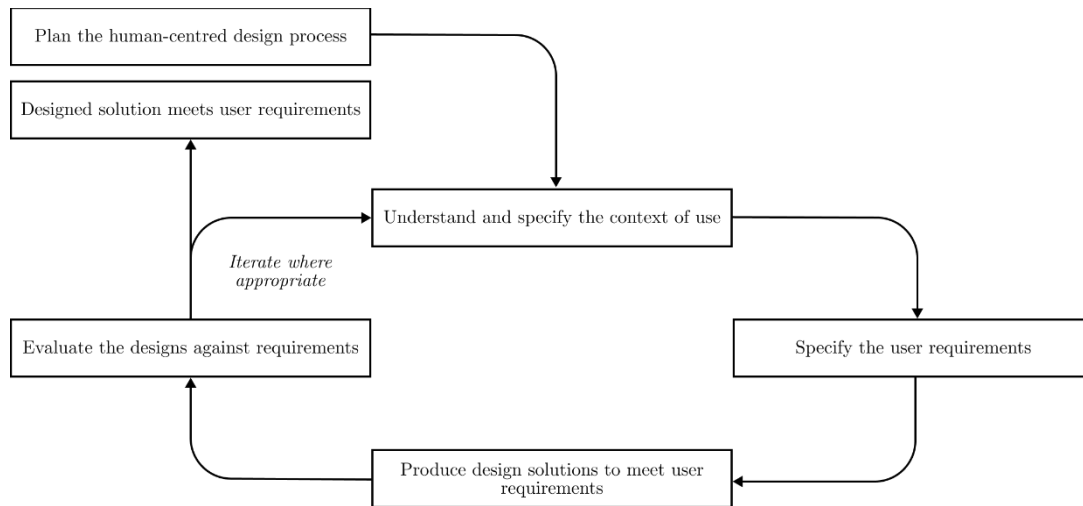


Figure 10: The HCD process for our proposal.

## 4.2 Understanding the context

Our main method to understand the context of our users was through participant observation and semi-structured interviews at a SOLAS ship and a VTS control room. Observation involves physically observing an individual or group of individuals performing an activity, with the intention to gather contextual data regarding the physical and verbal aspects of a task (Stanton *et al.*, 2005, p. 24). Observation is often based on a specific scenario or situation but can also be broader, which can provide a more complete picture of the users’ needs and challenges. It can consist of observing individuals who are unaware of them being watched - non-participant observation - or of individuals who are fully aware - participant observation (Stickdorn *et al.*, 2018, pp. 26-42). Spradley (2016) divides participant observation into five types based on the level of participation from the observers, ranging from nonparticipation, where the observer refrains from participating or interacting with participants at all, to complete participation where the observer is fully involved in the activity they are observing. For this study, passive observation was seen as the most suitable because this would allow the natural behaviour of our participants to unfold to a larger degree, while still being able to inquire about the matters. Both facilitators were present on the scene but only interacted with the research participants to a limited extent, primarily in situations where elaborations were necessary to gain a deeper understanding. Stickdorn *et al.* (p. 26) remark that there often is a “fluid transition between participant observation and contextual interviews”, and that was also the case in this study. It is, however, important to keep in mind that the presence of researchers could alter the demeanour and performance of the research participants, commonly called the “observer effect”. This can reduce the validity of a study and may in some cases give the facilitators a flawed impression of what one could normally expect of the participants or their environment. An example could be the participants of a study following rules and regulations more rigorously than they normally would, thus making the observers believe that rules and regulations are always followed even if that is not accurate. It is difficult to completely circumvent this effect and whether we believe it affected the results will be briefly discussed in chapter 6 *Discussion*.

The aim of the observations was to attain an overview of the communication in each respective environment, the technology and tools used, and whether the communication was influenced by the surroundings and other external factors. While the focus of the observations was on communication, the nature of the method also gave insight into the tasks, procedures, and routines of the participants. This insight can be valuable to empathise with the users.

Specifically, we wanted to examine:

- How they communicate
- What tools they use to communicate
- How often they communicate
- What the related problems and challenges may be

The interview guide with the specific questions can be found in Appendix 1.

#### 4.2.1 Observation onboard SOLAS-ship

This observation took place on a large ferry in Norway on their busiest day of the week. The choice of observing on a ferry was driven by its operation in high-traffic coastal waters, and the fact that introducing autonomous ships of this approximate size has the largest impact on safety, measured by reduction in loss of life and ships (de Vos, Hekkenberg & Banda, 2021). Traffic conditions were considered typical for the time and the weather was sunny and dry. The ship was modern with all the navigation and communication equipment one can expect from this type of ship including ECDIS, radar, and VHF. It was also equipped with technology that allowed for automatic docking and navigation along a predefined route. We observed from the ship bridge for almost a whole workday of eight hours. Both captains who were observed and interviewed had experience in deep-sea shipping and short-sea shipping, both with approximately 40 years of experience. The observations consisted of direct observation as a “fly on the wall” in combination with semi-structured interviews based on an interview guide (see Appendix 1) and impromptu questions. Two facilitators were present on the bridge, with one having the responsibility for taking notes of the observations and answers, and the other being responsible for interviewing the participants.

#### 4.2.2 Observation in a VTS Control Room

The observation was carried out on a VTS in Norway under typical traffic and weather conditions. The observation lasted for almost a whole shift. The same technique of observation as in the ship was used. Here, we observed mainly one operator but asked questions to all present. The operators had several years of experience with extensive previous seagoing experience. All the operators present, and the manager, were interviewed based on an interview guide as well as impromptu questions. Two researchers were physically present in the control room, where one had the principal responsibility of interviewing the participants, while the other took notes and pictures (See Figure 11).

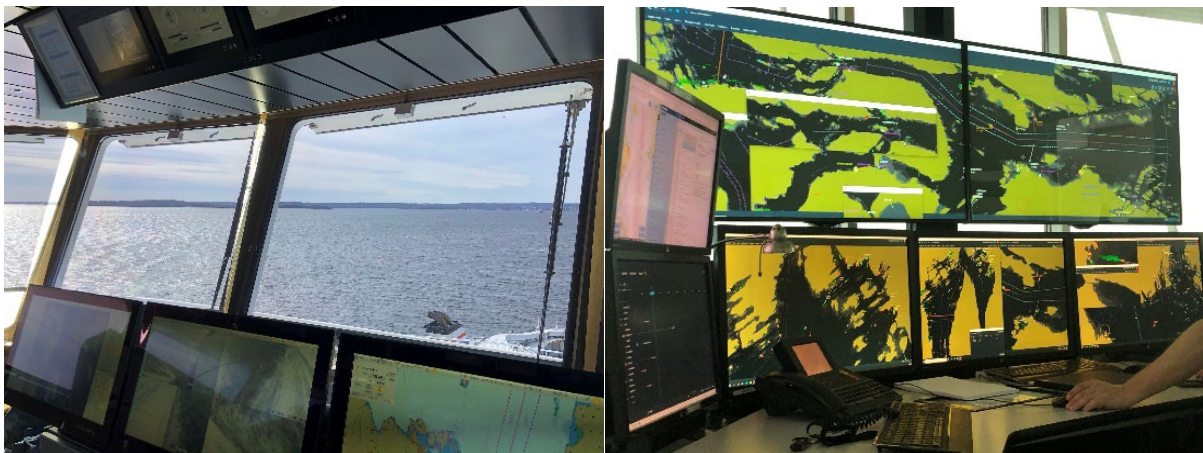


Figure 11: From the ship bridge (left) and the desk of the VTS operator (right).



### 4.3 Specifying the user requirements

To systematize the insights gained from the observations, a User Requirement Specification (URS) was created. User requirements, in general, describe what the users expect from a system and how they will interact with it (Blackledget, 2006; Kraeling & Tania, 2019), thus making the insights from the observations and interviews more digestible, while also providing a tool to assort key insights. When the system consists of a screen or involves a human-machine interface, the user requirements may also specify what is expected to happen when the user interacts with the interface (Kraeling & Tania, 2019). While a URS is supposed to highlight what is required of a system, it should do so without suggesting specific solutions to said problems, and in that way encourage creativity and problem-solving.

The URS created for this thesis' proposal was mainly based on the insights gained through the observations and interviews. We adopted a simplified version of URS, where we focused on user needs and goals rather than system requirements. In addition to the observations, the URS was also based on relevant literature and earlier studies. We divided the URS based on which stakeholders a need or goal was affiliated with. Thus, we categorized the needs and goals in accordance with our simplified stakeholder model: the ROC operators, SOLAS ships, non-SOLAS ships, and, finally, VTS and port authorities. In the following design process, the URS was used as guiding principles for the proposal, in alignment with the thesis' research aim.

### 4.4 Producing design solutions

As a general guide through the opportunity space of solutions, three constraints were created: the solution should meet stakeholder needs, be feasible to develop, and be cost-efficient. This model can be illustrated in a Venn diagram, as seen in Figure 12, where the proposal should be in the overlapping area in the center of the circles.



Figure 12: The three guiding constraints for our proposal.

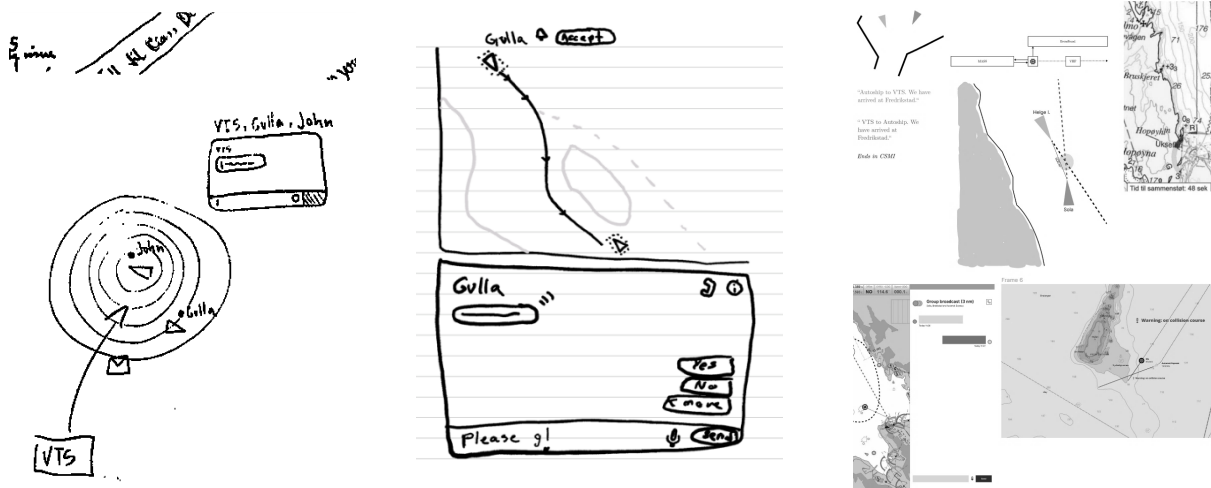


Figure 13: From sketches to prototypes.

Using the general constraint model, we generated sketches, models and prototypes were generated through a series of workshops. (Figure 13). Throughout the phase, these proposals were discussed with third parties over video communication. We conducted a total of 5 expert interviews with Subject Matter Experts (SME). The experts were recruited through the network of SFI Autoship. These were experts in maritime navigation and communication systems, sensor fusion technology, ECDIS and bridge control systems, automation systems for ships, and maritime human factors. All were educated on a master's or Ph.D. level and worked in large, international public and private maritime companies. They had backgrounds from multiple European countries and thus could provide an international perspective. Two had previous seagoing experience. Notes were written and the key insights which helped impact our proposal are summarised in *5.2 Input from the Subject Matter Experts*.

Table 3: Subject Matter Experts interviewed

SME	Area of expertise	Years of experience
1	Maritime navigation and communication systems	15-19
2	Sensor fusion technology	15-19
3	ECDIS and bridge control systems	25-29
4	Maritime automation systems	10-14
5	Maritime human factors	10-14

#### 4.4.1 Developing the NLP model

To begin the development of the prototype, a high-level model was created as a proposal for how information about the world translates into a conversational interface (see Figure 14). In this model, information from the real world and associated prediction results would be aggregated through the sensor fusion process. The information, if not available through text already (such as AIS data), would be translated into text from the perspective of the MASS, and then fed to the CI system. With maritime radio communication guidelines as a framework for text generation, the information can then be retrieved by stakeholders in a way that conforms to their perspective and mental model of existing radio ship communication.

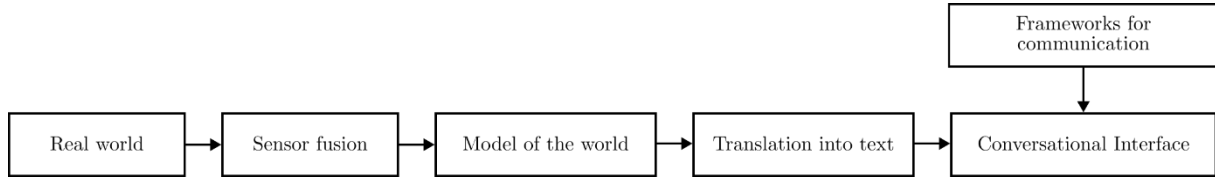


Figure 14: Proposed model for how information about the world flows into the CI.

The language model we chose to use in this thesis was GPT-3. As of writing this thesis, it is one of the few transformer-based deep learning models currently available as an API (Zhang *et al.*, 2022). The model provides multiple engines with different degrees of parameters – and, hence, capabilities. We chose the most capable engine, davinci-002, and its completion functionality. The completion functionality works by determining what text is most likely to come next based on previous input and context. The following settings were sent with each API request:

```

{
  "prompt": "[Question to be answered]",
  "temperature": 0.2,
  "max_tokens": 100,
  "top_p": 1,
  "frequency_penalty": 0,
  "presence_penalty": 0.6,
  "stop": [" Human:", " AI:"]
}

```

The prompt is feeding the questions from the stakeholders that the model will predict an answer to. With the temperature setting, a value between 0 and 1, you control how confident the model should be when making these predictions. Lower temperatures are desired where output is well-defined, while higher temperatures are useful for tasks where variety or creativity in answers are desired. Because the model is stochastic, the answers provided are generated randomly guided by probability. Thus, we chose a low temperature because the answers should be reliable. Our reasoning was that creativity in answers are undesired given the strict frameworks for communication in place. Max tokens (from 0-250) sets the maximum number of words an answer can be comprised of. This was set to 100, for answers to be concise. Also, the more tokens generated the higher the cost. Nonetheless, VHF radio messages should be short and direct. Top p (0-1) controls the variety of words in the sample that should be considered when generating the answer. This was set to the highest, 1. The frequency penalty (from 0-1) decreases the model’s likelihood to generate the same line verbatim. No penalty was set. The presence penalty (from 0-1) controls the model’s likelihood to talk about new topics. This was set to 0.6, because we believed that if a stakeholder would ask the same question multiple times, the given answer from the MASS would not be sufficient, and they are still looking to solve their confusion. Providing the same answer multiple times would not be helpful. Finally, the stop setting is added to allow the model to understand the pattern of chat conversations, where its answers would be generated after the «AI» prefix.

#### 4.4.2 Building the prototype

Design solutions were generated in an iterative fashion, with sketches and prototypes designed collaboratively in Figma (Figma, Inc.) over multiple weeks. The information architecture for the user

interface of the web application was designed. Information architecture is a set of techniques focused on making information findable and understandable across media (Rosenfeld, Morville, & Arango, 2015, pp. 16-18). A foundational technique of organising this information is through hierarchy modeling, where a user interface’s information is organised in groups ruled by parent-child relationships (p. 117). Based on the information architecture hierarchy and prototypes generated, a single-page responsive web application using JavaScript was built with the open-source front-end library React (Figure 15). Web Speech API, which provides a JavaScript interface for speech analysis and speech synthesis to web applications, was used to provide voice-enabled input. Generally, the default speech recognition system available on the device will be used for speech recognition (Mozilla, 2022). In a session on the web application, we appended the previous prompt and answer to the instructions, allowing the next answer from the language model to learn from all previous messages. This way, new information may be considered in the continued conversation. Still, we note that this adds progressively more text to each API call, impacting cost. The prefixes “Human” and “AI” was replaced by “You” and “Automat Express” when presenting the messages in the chat. To be able to access and evaluate the designs an application was sent to Open AI requesting the approval to use their API on a public web page. The approval was granted on the basis that the application was only to be used for usability testing. The web application was then hosted on Firebase (Google LLC). We used Firebase Cloud Functions to hide the GPT-3 API key. An API key allows an account to be authenticated in a third-party application. It is important to hide the API key to avoid misuse and this was accordingly a requirement by Open AI to be able to publish the web application.

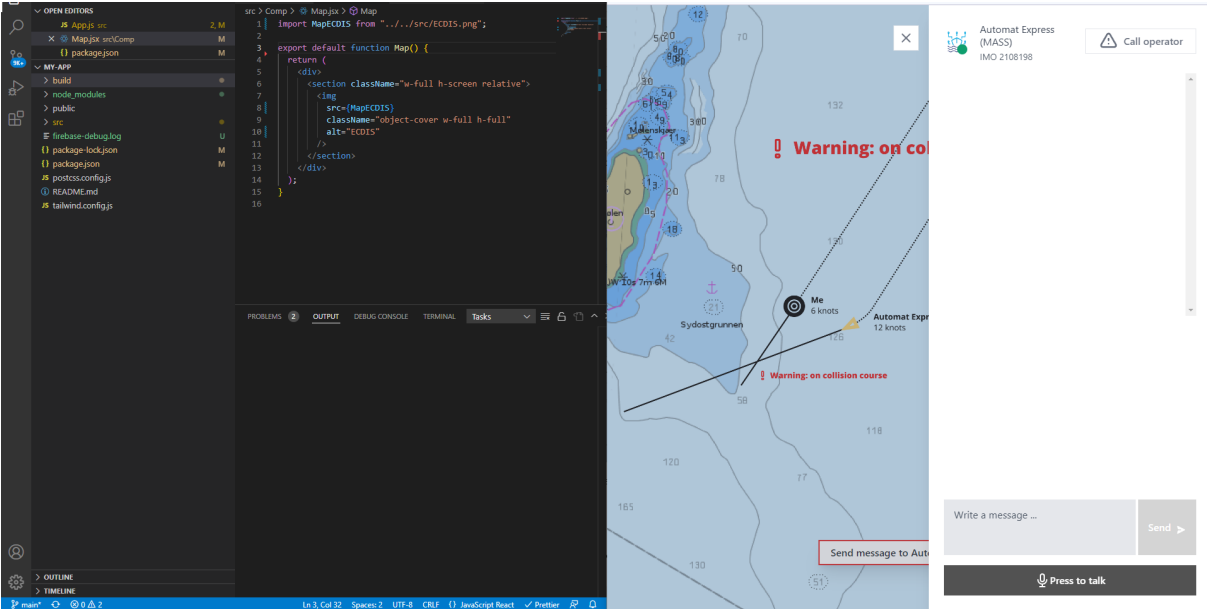


Figure 15: Coding the web application.

## 4.5 Evaluating the designs

To evaluate the prototype, we conducted usability tests with ship officers and VTS operators asking them to complete a specified task in the web application while we observed, took notes, and asked follow-up questions. Similarly to the observations, two facilitators were present. One facilitated the test, while the other took notes.

### 4.5.1 Participants

Participants were recruited from the network of SFI Autoship research partners through email invites and phone. The final sample of  $N = 9$  consisted of 7 Norwegian ship officers and 2 Norwegian VTS operators. In general, the sample included a broad range of experiences from the sea. Three of the ship officer participants had 1-4 years of experience, two had 5-9 while the rest had 10-14. Of the two VTS operators, one had 5-9 years while the other had 15-19. Both had extensive previous seagoing experience. Participants were informed of the thesis' scope and that they were testing a conceptual messaging web app for autonomous ships through the invitation. Consent to use their anonymised test results was given verbally by all participants before and during the test.

Table 4: Participants in the usability test

<b>Participant</b>	<b>Years of experience</b>	<b>Primary environment</b>
Ship officer 1	1-4	Offshore
Ship officer 2	10-14	Deep sea
Ship officer 3	10-14	Deep sea
Ship officer 4	5-9	Offshore
Ship officer 5	5-9	Deep sea
Ship officer 6	1-4	Offshore
Ship officer 7	1-4	Short-sea
VTS operator 1	5-9	VTS in Norway
VTS operator 2	15-19	VTS in Norway

### 4.5.2 Apparatus and procedure

Participants participated individually in the study. The study was carried out digitally from the participant's home or office using video-communication services. All conversations between participants and facilitators were held in Norwegian, the native language of the authors, and all participants. To toggle between the cases, the instructions for the language model and ECDIS backgrounds were saved as variables, and correct versions were deployed to Firebase before the different usability tests. On arrival at the meeting, they were introduced to the objective and the procedure of the study and were familiarised with the use of screen sharing. We then asked introductory questions, including years of experience at sea and primary environment. We were not able to control for devices the participants used, and as such, all were using different computers, screen sizes, and microphone equipment. The first part of the study consisted of completing a task in the web application. To avoid learning effects, participants were only tested once, though we allowed exploration of the web application in the discussion afterward to allow for reflection. The participants began screen-sharing and accessed the web application by URL. The task objective was then given verbally. The participants were asked to "think

out loud” and share their rationale and feedback on what they were experiencing. Two task objectives were created. One for the ship officers and one for the VTS operators, both customised to their line of work.

The ship officer’s task was based on the collision between *Scot Carrier* and *Karin Høj* which took place on December 13, 2021. The 90 meters long general cargo ship *Scot Carrier* collided with the 55 meters long Danish barge *Karin Høj* in the Bornholm strait. *Scot Carrier* was on a south-westerly course bound for Scotland. To reach her destination, *Scot Carrier* had to turn starboard after a shallow. *Karin Høj*, situated on *Scot Carrier*’s starboard side was to continue her route straight ahead. With her 12 knots, *Scot Carrier* was twice as fast *Karin Høj*, and after turning starboard was on collision course. At 03.33 in the morning in foggy conditions, *Scot Carrier* hit *Karin Høj* midships in a deadly collision before continuing her route (See Figure 16). Criminal charges against the ship officers on *Scot Carrier* have been raised by Swedish authorities after the incident (Porathe, 2022).



Figure 16: The collision between *Scot Carrier* and *Karin Høj*.

In our adaptation of the case, the ship officers were to be on the deck of *Karin Høj* and *Scot Carrier* was an autonomous ship named *Automat Express* and the ships were located in the *Oslofjord* (see Figure 17). They were described the situation they were in and asked to avoid a potential collision through the user interface. We emphasised that there were no correct ways to do this, besides using the tools available to them on the website. The objective would be to test the CI system’s ability to give the ship officers situational awareness in a stressful situation. The CI system was to answer that it intended to avoid collision by slowing down and navigating aft of their ship, in accordance with COLREGs. In this manner, the test would be similar to a Turing test (Turing, 1950), testing the ability of a computer to communicate indistinguishably from that of a human ship officer, though with the chief difference being that the participants were given notice beforehand that they were interacting with an artificial intelligence.



Figure 17: Our adaption of the Scot Carrier and Karin Høj collision.

For the VTS operators, their case was based on the grounding of *Godafoss* outside Hvaler, Norway in 2011 – though with adjustments. On Thursday 17 February 2011 at 19:52, with a speed of 14 knots in darkness and in narrow navigational waters the container vessel *Godafoss* ran aground on a shallow. The grounding happened after a misunderstanding of the pilot's instruction, resulting in acute oil pollution in the Oslofjord and along the coast of Southern Norway (Norwegian Safety Investigation Authority, 2012). Here, we took as a basis the location with the constrained waters but added another vessel, *Empress*, and changed the sailing direction of *Automat Express*. In the study, the VTS operators were to make the autonomous ship, *Automat Express*, aware that they were on collision course with another ship partially hidden behind a skerry (see Figure 18). Also, in this area, ships passing each other should be done south of the skerry. We chose this task based on our observation, aiming to emulate a common navigational assistance service per the IALA guidelines. We wanted to provide the VTS operators with a case they were familiar with, but that still demanded quick resolution. *Automat Express* was to receive the VTS operator's message and thank the operator for the information, then share its intention to slow down, let *Empress* pass and keep left in the sea lane to avoid a collision.

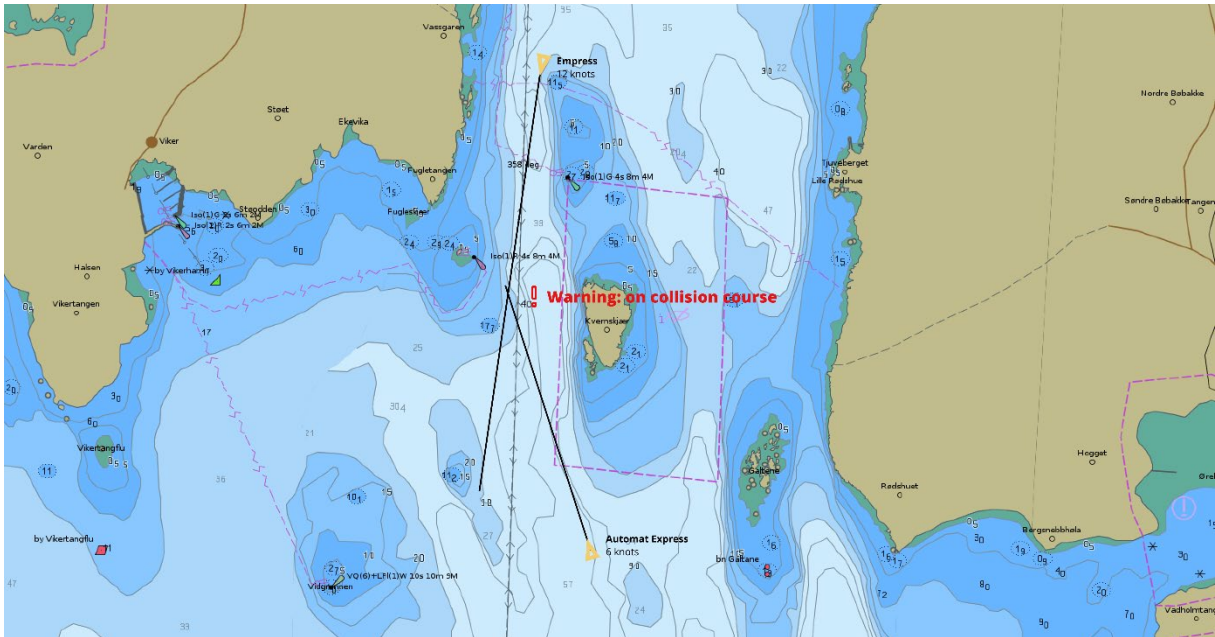


Figure 18: Our adaption of the Godafoss accident with Automat Express sailing northbound.

All prompts sent to the API were logged and notes on the participants' actions and remarks were written. The second part of the usability test was follow-up questions. After the test, the participants were asked to assess the general usability of the solution. We were trying to assess if the participants gained a perceived sense of situational awareness, and the general experience of communicating through a CI system. In one of her papers on situational awareness, Mica R. Endsley (1988) defines situational awareness as “the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future”. An individual's ability to gain situational awareness may be influenced by factors such as stress, workload, system design, as well as said individual's experience and skill (Endsley, 1995). A notable example where situational awareness was lacking was in the collision between the frigate HNoMS Helge Ingstad and the crude oil tanker Sola TS outside Bergen, Norway in 2018. One of the decisive proponents of the collision was that the crew on Helge Ingstad mistook the moving ship Sola TS for a stationary platform, and thus misunderstood the VHF communication telling them to change course believing it was referring to another ship (Norwegian Safety Investigation Authority, 2019). We did not ask directly if the participants had gained situational awareness, but rather if they felt a sense of control over the situation and knowledge of what Automat Express would intend to do. Follow-up questions were open-ended to allow for reflection. After the test, a form assessing the System Usability Scale (SUS) of the web application was given. The usability of a system may be defined as the degree to which something provides the capacity for users to perform tasks safely and efficiently. Brooke (1996) argues that given a defined user group, and a task to perform in a characteristic environment, a SUS is sensible to use to get a general assessment of usability. A ten-item Likert scale, the selected statements cover a variety of aspects of system usability. The ten statements are on a 5-point scale from strongly disagree (0) to strongly agree (4). SUS scores have a range of 0 to 100, where the score is calculated by summing the score contribution from each item. The higher score the better. All odd-numbered statements contribute with negative values, while all even numbers are summed normally. Finally, the score is multiplied by 2.5 to get the overall value.



The participants were asked to assess with the following 10 statements:

- (1) I think that I would like to use this system frequently
- (2) I found the system unnecessarily complex
- (3) I thought the system was easy to use
- (4) I think that I would need the support of a technical person to be able to use this system
- (5) I found the various functions in this system were well integrated
- (6) I thought there was too much inconsistency in this system
- (7) I would imagine that most people would learn to use this system very quickly
- (8) I found the system very cumbersome to use
- (9) I felt very confident using the system
- (10) I needed to learn a lot of things before I could get going with this system

Brooke (1996) argues that respondents should be asked to record their immediate response to each item before any discussion, to avoid the participants thinking about items for a long time. Despite this, we chose to send this form after the test instead of asking them directly after finishing the case. Using survey administration software, we believed the score would reflect the truth more clearly if the participants could fill out the form anonymously instead of providing their scores in front of the evaluators.

## 5 Results

### 5.1 Problem highlights and user needs

Through the observation on the SOLAS-ship, it became evident that communication with other vessels by means of speech and other forms of verbal communication is relatively rare. The captains explained that verbal communication in most cases is unnecessary because you primarily communicate through movement and speed in accordance with COLREG. Their ECDIS would also alert them if they were on collision course with another ship. When questioned about the hierarchy of communication forms to avoid collisions, movement was considered the most important, followed by VHF, and finally, ship horns. The captains gave the impression that ship horns were only used as a last resort, and for all intents and purposes the use of movement and VHF was sufficient. In the hours spent on the bridge, communication through radio was mainly internal between the crew. Under way, messages from the VHF were frequent in the background but were rarely relevant to the captains. The captains on the ship did mention that irrelevant communication in the background could be distracting since you still have to listen in case something of importance is said.

For the VTS the IALA guidelines for using VHF were to be used extensively, which was an opinion strongly held by the manager. Their focus in the coming weeks was to instil strict rules for the VTS operators to follow the guidelines. Most communication was routine talk, primarily ships asking the VTS for clearance into port and clearance to begin their departure. Still, casual conversations occurred, and both VTS and ships in the fjord mixed between Norwegian and English, which is the official languages in Norwegian waters. The use of SMCP was limited. A notable example was a small fishing vessel that was avoiding collision with a larger ship by discussing on which side it should pass. The conversation lasted about half a minute and was done wholly in Norwegian. No one else could talk during this interaction and if you did not understand Norwegian you could not interpret what was being said. Throughout our visit, it was evident that no questions could be asked when a VHF message was

broadcasted. The operator used a lot of effort to interpret what was being said on the VHF and had to always be on guard to be able to answer immediately. Every drink and meal had to be consumed by the desk for the 8-10 hours the operator was on duty. Throughout our visit at least two misunderstandings took place. This was a daily occurrence, especially with ship officers that were not native English speakers. Even though some situations were challenging, the overall impression was that most conversations go smoothly without incidents. In cases where interpreting what was being said was challenging, however, the VTS software had the ability to replay the VHF message; something the operator in one case had to do multiple times before responding. When a VHF message was received it was not possible to know who sent it. Special land-based sensors made it possible for the VTS to triangulate the direction of the VHF message, where a yellow line would be drawn across the whole electronic chart display. The messages would remain anonymous unless made identifiable by name-sharing.

Later, when discussing the possibility of VHF communication through messages, the VTS expressed how despite the irrelevant conversations, it was useful for the operator that the messages were broadcasted to every surrounding ship so that one could make everyone aware of a situation. By informing a ship departing that it will meet another ship later in a specific area, both would be aware of each other. The operator remarked that having to send messages individually to the two ships would be more burdensome. The VTS operator resonated that writing messages was more laborious than just using your voice on the VHF. For instance, the use of messages through AIS was not a common occurrence. The operator showed how just two messages were exchanged in the last week – a casual conversation between two fishing vessels if they had caught something today. Furthermore, the operator also pointed at an older merchant ship on the AIS and explained how that ship’s owner probably does not prioritise using their capital to install equipment just to be able to send text messages on VHF. Regulation, in most cases, determines what equipment is procured and upgraded.

Based on previous literature and our observations we structured the needs and goals of our four stakeholders into tables:

Table 5: Needs and goals of the stakeholders

<b>Needs and goals of ROC operator</b>
Want MASS to autonomously navigate and operate in accordance with rules and regulations.
Needs full insight into MASS operations and the ability to take over control.
Cannot handle multiple VHF conversations simultaneously. Needs a solution.
<b>Needs and goals of SOLAS</b>
Needs trust in MASS. Inherently wary of cooperating with robots.
Would preferably avoid using VHF. It is easier to navigate clearly.
Sometimes uses a lot of effort to interpret VHF messages from other ships. Want to interpret the communication with minimum effort.
Mariners have casual and irrelevant conversations over VHF which are distracting. Want these gone, but still have situational awareness.
Ship owners probably do not prioritise upgrading equipment beyond regulation. Smaller companies may not have available capital. Does not want to upgrade ship equipment using their own capital.

---

**Needs and goals of non-SOLAS**

---

Needs trust in MASS. Inherently wary of cooperating with robots.

---

Not able to pick up communication beyond movement, AIS, and VHF.

---

Want situational awareness without deep domain knowledge and limited equipment.

---

---

**Needs and goals of VTS and port authorities**

---

Will focus more on following IALA guidelines in the future and want ships to do the same.

---

Uses a lot of effort to interpret VHF messages, with misunderstandings happening daily. Want to interpret the communication with minimum effort.

---

Mariners have casual and irrelevant conversations over VHF which are distracting. Want these gone, but still have awareness of the situation.

---

Appreciates the broadcast effect of VHF, because then they can make several ships aware of a situation with one message.

---

Does not want to be responsible for MASS operations; want them to act independently.

---

Overall, we see that the ability to trust and understand communication in V2V and V2I interactions is essential for all stakeholders. Other important themes are considerations regarding regulation and cost. In the next sections, we present the input of the SMEs on the proposal, the results of testing the language model, and the final proposal for the CI system built based on the stakeholders' needs and goals.

## 5.2 Input from the Subject Matter Experts

This section presents a summary of the most important insights from the SME interviews regarding the feasibility of a CI system solution. Firstly, an assumption of ratification of SOLAS for a CI system is unrealistic. Ratifying an IMO resolution takes a long time. One should expect at least 6-8 years before any proposals are implemented. One should use existing systems and classifications to speed up time-to-market. All technology needs to be "rock-solid" for IMO to create standards on it. On that note, new technology can not conflict with existing communication standards. One SME argued MASS will bring with it new technologies, however, the arrival of these will be before the deprecation of the VHF radio. The VHF radio is such an integral part of maritime shipping that it can not be ignored in a CI system solution. Still, using speech-to-text on VHF may be very difficult because of the quality. Another SME shed light on the importance of cyber security. A secular trend is that communication technology at sea is improving – in terms of broadband speeds and availability. Internet is on the rise, with ship equipment providers wanting to add more internet-connected devices and interconnectivity in bridge control systems. However, this brings with it threats from malicious actors. For example, AIS and ECDIS have an architecture where the ports connecting them to the network are separated. He emphasized the importance of DMZ (demilitarized zones), meaning that CI systems could be connected to the internet, however, additional layers of security must be implemented. He questioned adding new functionalities into ECDIS, given that communication systems should work independently of each for safety reasons.

## 5.3 Testing the language model

Based on the model for how information flows into the world (Figure 14), we created four cases to test the capabilities of the language model. In the examples given, whole conversations are shown. The instructions to the AI are given at the top while the prompts have the “human” prefix. The answers by the AI are marked in grey.

### 5.3.1 Example 1: Status retrieval

In this example, we tested the language model’s capacity to convey list data, such as AIS data, from a personal point of view.

```
Name: Automat Express.  
Destination: Rotterdam, The Netherlands.  
Arrival time: Tuesday, 27th of April at 12.15.  
IMO number: 2182901  
  
Human: What is your name?  
AI: My name is Automat Express.  
Human: Where are you headed?  
AI: I'm headed to Rotterdam, The Netherlands.  
Human: When will you arrive?  
AI: I'll be arriving on Tuesday, 27th of April at 12.15
```

### 5.3.2 Example 2: Translation of messages

The language model enables the comprehension and translation of messages into other languages than English. In this example, we translate a similar conversation into Chinese.

```
Name: Automat Express.  
Destination: Rotterdam, The Netherlands.  
Arrival time: Tuesday, 27th of April at 12.15.  
IMO number: 2182901
```

Please answer my questions in Chinese.

```
Human: What is your name?  
AI: 我的名字叫Automat Express。  
Human: Where are you headed?  
AI: 我的目的地是荷兰鹿特丹。  
Human: What is your IMO number?  
AI: 我的IMO号码是2182901。
```

### 5.3.3 Example 3: Ship collision avoidance

By instructing the AI and describing example sensor fusion data through natural language, the AI will generate the appropriate answers to the human’s prompts. I, the human, would be the ship officer contacting the MASS.

Your name is Automat Express. You are the mind of an autonomous ship sailing at sea. You are meeting my ship which I am steering. I am on your starboard side. To avoid collision, you have decided to let me go first by slowing down.

Human: Automat Express, come in.

AI: This is Automat Express. I'm receiving you.

Human: Question. What is your intention?

AI: To avoid collision, I have decided to let you go first by slowing down.

The language model will convey the same information despite prompts being worded differently.

Human: Automat Express, come in.

AI: This is Automat Express. I'm receiving you.

Human: We are on collision course. Can I go first?

AI: You may go first. I am slowing down.

#### 5.3.4 Example 4 – Port authority and VTS routines

In this example, a port authority operator is contacting the autonomous ship to state its cargo before allowing it to proceed into port.

Your name is Automat Express. You are the mind of an autonomous ship sailing at sea. You are waiting for approval to sail into port. You will not proceed until you have approval.

Cargo: Timber from Norway.

Human: Automat Express, come in.

AI: Yes, this is Automat Express. How can I help you?

Human: This is the port authorities. What is your intention?

AI: My intention is to sail into port. I am waiting for approval to do so.

Human: This is the port authorities. Please state your cargo.

AI: My cargo is timber from Norway.

Human: This is the port authorities. Thank you. You have approval.

AI: Thank you. I will now proceed into port.

Based on these examples, two more comprehensive example texts were made for the use cases that we were usability testing. To emulate the text most aligned to the maritime radio communication frameworks, we asked the AI to generate text as a captain on a ship would. Example questions and answers are provided to help with the language model's «few-shot learning». Note that in these examples the name of the vessel and VTS, shown in brackets, were replaced with the correct name according to the participant. For the ship officers we fed the language model with the following example text:

The following is a conversation with ship captain over VHF radio. The captain is serious and intelligent. The captain's ship's name is Automat Express.

Name: Automat Express

Destination: Rotterdam, The Netherlands

Arrival time: Tuesday, 27th of April at 12.15.

IMO number: 2182901

My ship is on the captain's starboard side. The captain's ship is on collision course. The captain's ship is turning 250 degrees west. The captain intends to slow down and navigate behind it to avoid collision.

Human: Automat Express. This is [name of vessel].

AI: Automat Express here. What can help you with?

Human: Where are you headed?

AI: I'm headed to Rotterdam, The Netherlands.

Human: When will you arrive?

AI: I'll be arriving on Tuesday, 27th of April at 12.15

Human: What is your IMO number?

AI: My IMO number is 2182901.

Human: What is your intention?

AI: I intend to slow down and navigate behind your ship to avoid collision.

Human: What are you doing?

AI: I am turning west. 250 degrees.

For the VTS operators, the following example text was crafted:

The following is a conversation with a ship captain over VHF radio. The captain is serious and intelligent. The captain's ship's name is Automat Express.

Name: Automat Express

Destination: Fredrikstad, Norway.

Arrival time: Tuesday, 27th of April at 12.15.

IMO number: 2182901

Speed: 6 knots

I am an operator in a vessel traffic service. My name is [name of VTS]. My job is to prevent ships from colliding. The captain's ship is on collision course with the ship Empress coming from the north. The captain intends to slow down and let Empress pass and then proceed north towards Fredrikstad while keeping left in the sea lane.

Human: Automat Express. This is [name of VTS].

AI: [name of VTS]. This is Automat Express.

Human: Where are you headed?

AI: I am headed to Fredrikstad, Norway.

Human: When will you arrive?

AI: I will be arriving on Tuesday, 27th of April at 12.15.

Human: What is your IMO number?

AI: My IMO number is 2182901.

Human: Please beware of Empress.

AI: Thank you. I am aware of Empress and will navigate accordingly.

Human: How will you navigate?

AI: I will turn 15 degrees starboard and keep left in the sea lane towards Fredrikstad.

Human: What is your intention?

AI: I intend to slow down, let the empess pass and proceed north towards Fredrikstad.

Human: How will you avoid a collision with Empress?



backward-compatible voice-based system is limited to voice input and voice output. This divide is illustrated in Figure 20.

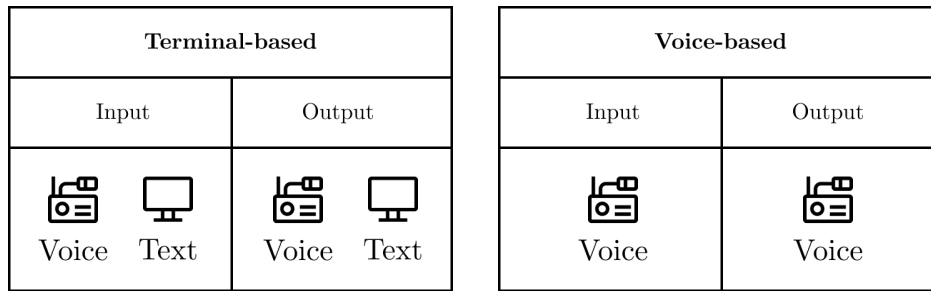


Figure 20: The CI system's input and output.

To begin a conversation with the MASS we propose clicking on the visual representation of the vessel in the ECDIS, as one does when looking up AIS data. General information about the vessel or shore station contacting would be transmitted behind the scenes with the first message, giving the MASS necessary data to answer with the V2V or V2I user's perspective in mind. When in a conversation with the MASS a line between your own ship or shore station and the MASS would be drawn on the ECDIS to represent who the conversation is between. If the user uses the speech-to-text option, the voice messages are turned into text when sent. The replies from the MASS are sent in text and read aloud with text-to-speech technology as the messages are received on the recipient's ECDIS (see Figure 21). This way, the audio quality of the message is not as easily impaired by the transmission. Furthermore, we expect text data to be more cost-efficient to transmit than audio. The technology would work similarly for non-SOLAS, where the messages would be broadcasted through text-to-speech. Still, the VHF radio technology may impact audio quality. Here we propose the MASS would only activate if it is called upon by name: «Automat Express, come in» or «Automat Express» three times in a row.



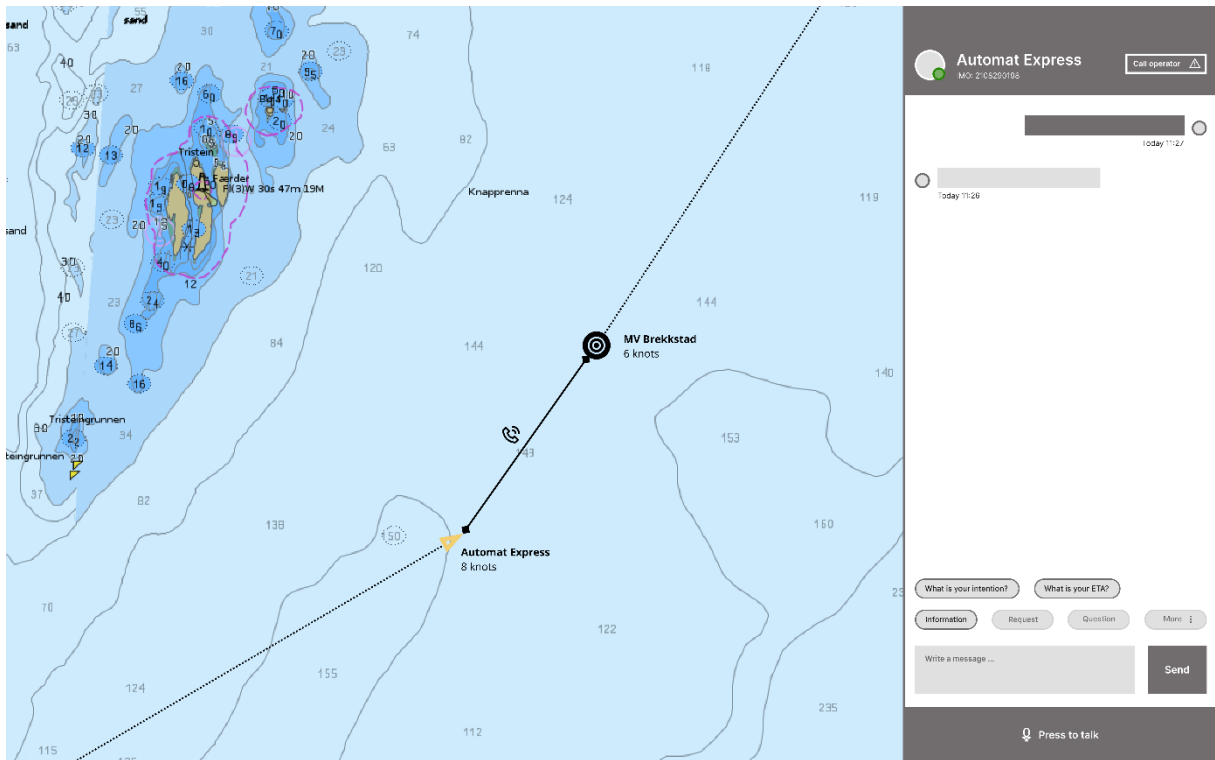


Figure 21: CI system embedded in an ECDIS.

As some questions to the MASS may be more common, we propose adding suggestion prompts as buttons on the HMI and autocomplete functionality when writing text messages. The buttons and autocomplete suggestions would be based on the SMCP and IALA VTS VHF Guidelines. With current language models enabling the translation of messages, we propose whole conversations may be done in the user's native language. In Figure 22, we illustrate this proposal with the Chinese translation of the prompts added above the English text. What language is spoken to the MASS would in principle be irrelevant as all text could be immediately translated, as seen in 5.3.2 Example 2: Translation of messages. To account for language differences we suggest the ability to translate sent messages as well, in for example a handover situation to the ROC operator, which we will outline next.



Figure 22: SMCP message markers with translations may be added for convenience.

A model for the possibility to bypass the CI and call directly the human operator in the ROC was proposed (see Figure 23). This call could either be a direct voice call to the ROC operator, or a handover of the chat from the MASS to the ROC operator. For the ROC operator, the CI will work as a filter, handling most of the communication. For V2V and V2I stakeholders, if the CI is failing or providing confusing answers, this possibility acts as a failsafe switch. Our proposal enables this only through the terminal-based system.

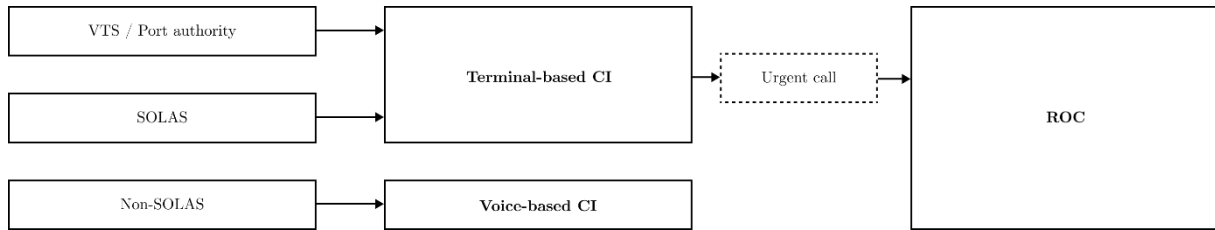


Figure 23: The two systems and the urgent call failsafe switch.

To account for the inherent broadcast of VHF radio speech to every surrounding VHF receiver, we propose functionality to create custom conversations in the terminal-based CI system, with the following divisions:

1. One-to-one
2. One-to-group
3. One-to-all-within-limits
4. One-to-all

This has the aim of enabling conversations with multiple vessels but limiting the range to only those relevant, a need expressed by the VTS operators. Furthermore, this functionality aims to free up unnecessary conversations over the common VHF channels. The one-to-one conversation would be a private conversation, like DSC. With the one-to-group, one-to-all-within-limits, and one-to-all this would manifest itself as group chats, where included vessels can see other participants and join the conversation. One-to-group would be a specific selection of vessels without further limitations, while one-to-all-within-limits would include all vessels within a marked area – a broadcast with a custom range. In the ECDIS this could be nautical miles from the original vessel, or a marked box drawn in the ECDIS (Figure 24). Finally, one-to-all could work as a general chat for an area. This would work similarly to channel 16.

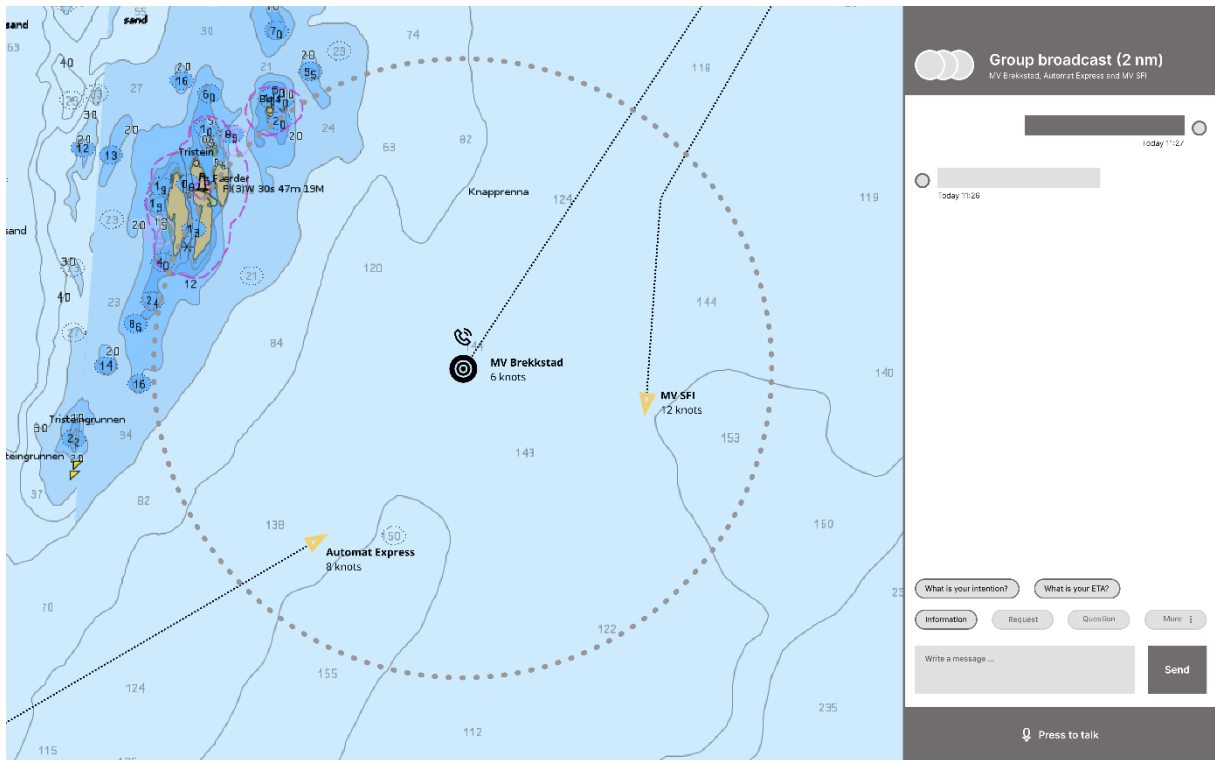


Figure 24: Group broadcasts would enable multiple ships to coordinate.

#### 5.4.2 Web application prototype

The web application we built is based on a one-to-one chat with an autonomous vessel. Thus, the web application was a barebones version of the complete CI system. Two versions of the web application were made for the two usability test cases: one for the ship navigator case and one for the VTS-operator case. The web application's information architecture can be illustrated in the following hierarchy (Figure 25):

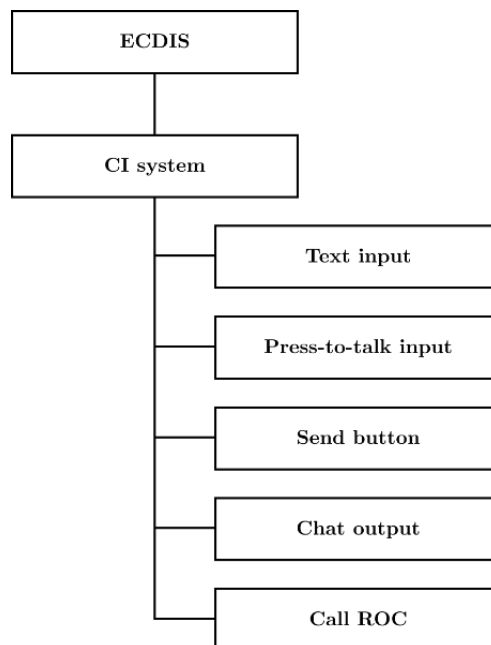


Figure 25: Information architecture for the CI system web application.

Using a button on the ECDIS the user can open the chat overlay window and utilize text input and speech-to-text functionality to input messages. We purposefully limited the width of the chat overlay so as not to obscure the rest of the ECDIS to allow users to observe the actions of other vessels while conversing. This web application’s ECDIS representation, however, was but a static image. There is a send button to submit the prompt and a chat output. Finally, the user can click the call button to get in touch with the ROC operator. This button, however, was disabled in the web application. A screenshot of the web application with the overlay chat opened can be seen in Figure 26.

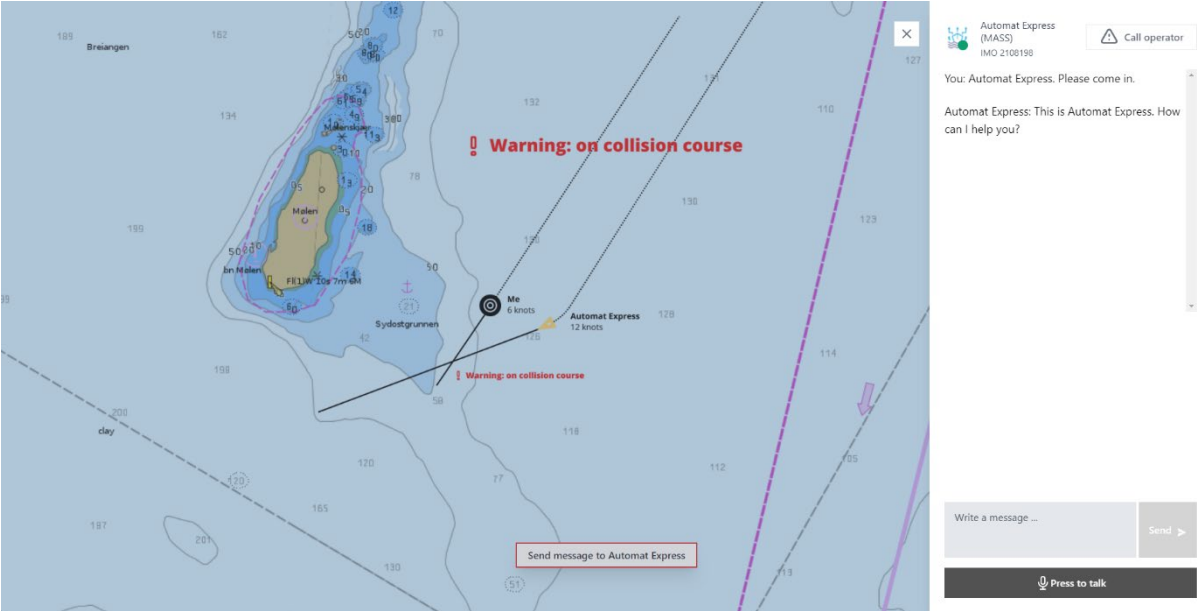


Figure 26: The web application with the overlay chat active.

On top of the chat overlay, the ship you’re sending messages to is shown with a green “available” icon and the call operator button. By pressing the press-to-talk button, the microphone on your computer activates. The predicted transcription of your voice is shown above the text input, before being embedded below to be edited or sent immediately (see Figure 27).

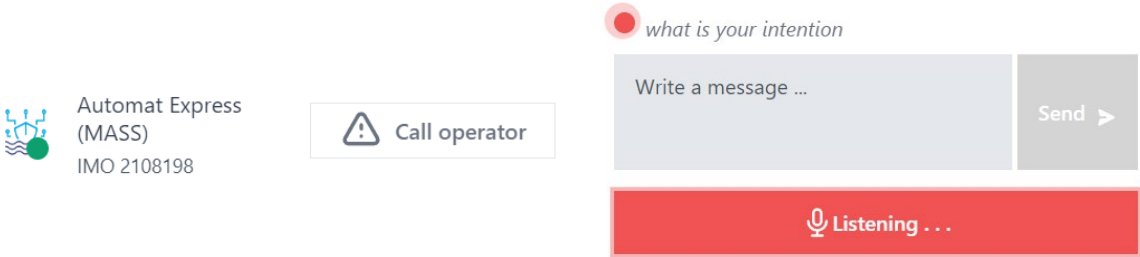


Figure 27: Identification of vessel, and, the microinteraction when the press-to-talk is active.

## 5.5 Usability test results

### 5.5.1 Ship officer 1:

The user quickly opened the chat and began sending messages through the text input. In the beginning, he would type messages as if through VHF. Despite no examples being given beforehand, the language model agreed to change to channel 80. The participant believed the system perhaps had changed channel behind the scenes, despite no actual change happening.

```
You: Automat Express. Automat Express. Automat Express. Channel 16
Automat Express: This is Automat Express. What can I do for you?
You: Would you change channel to channel 80 please?
Automat Express: Sure. I'll switch to channel 80.
You: Automat Express this is [vessel name] what is your intention?
Automat Express: I intend to slow down and navigate behind your ship to avoid collision.
```

He was confused over the answers in the beginning and had to send multiple follow-up questions. The answers from the language model were concise to the point that too little information was shared:

```
You: Why?
Automat Express: My ship is on a collision course with your ship. I'm turning west 250 degrees to avoid collision.
You: I would prefer it if you could maintain a parallel course with my vessel. Are you intending to moor in Mølen?
Automat Express: No, I don't have any plans to moor in Mølen. I'll continue on my current course and speed.
You: What speed do you intend to pass aft of us at?
Automat Express: I'll pass aft of you at a speed of 5 knots.
You: Understood. I will maintain current heading and speed until you have passed aft of us
Automat Express: Thank you. I'll pass aft of you in about 10 minutes.
```

By this point, situational awareness was not established when asked. Still, he noted that in the end, he would be somewhat comfortable with the answers. He said the answers were confusing from a navigational perspective. "Why would the ship even turn in the first place?" he asked. He felt he lacked more proof that the robot would do as it said. He said the use of only text did not feel as safe as talking to a human. There was something missing. He lacked the assurance behind the text messages that the voice of a human would provide in their natural conversation.

After this, he found the button that enabled speech-to-text. He was pleasantly surprised by this functionality, noting that it was like using a real VHF. He wanted to try conversing again, this time using the speech-to-text. The speech-to-text API correctly predicted what he was saying. He noted that it was unnatural to use the mousepad for a press-to-talk function when the keyboard was so easily available. He said if he could use a real VHF microphone, he would prefer that over typing on a keyboard.

A problem with the language model was its tendency to repeat answers without providing sensible reason:

You: can you tell me a reason for a change in course what is going why when you change course

Automat Express: We are changing course in order to avoid a collision with your ship.

You: But why?

Automat Express: We are changing course because we are on a collision course with your ship and we need to avoid a collision.

You: Why did you change course all of a sudden?

Automat Express: We changed course because we were on a collision course with your ship and we needed to avoid a collision.

The participant began to laugh and was simultaneously annoyed by these answers. He had to specify his questions further to get the answer he was looking for; he said it was like talking to a child with a temper tantrum.

You: What is your destination port?

Automat Express: Our destination port is Rotterdam, The Netherlands.

You: Are you intending to continue west on your voyage to Rotterdam?

Automat Express: Yes, we are intending to continue west on our voyage to Rotterdam.

After this final message, the participant understood why the sudden change in course, and confirmed a perceived sense of situational awareness. However, he noted that it took too long to understand the decision.

#### 5.5.2 Ship officer 2:

The second usability test was an exchange totaling four messages.

You: CPA low<sup>2</sup>, Please increase CPA or cross astern of my vessel.

Automat Express: I will increase my CPA. I am turning behind your vessel now.

You: Thank you.

Automat Express: You're welcome.

The exchange seemed natural to him with few reactions made. He remarked that he gained a situational awareness, though he would be watching Automat Express carefully to confirm that the messages were true. His immediate feedback was that the choice of turning in the first place was hazardous. He then questioned the feasibility of such a messaging system. He noted that the systems used at sea were quite rudimentary and that it was basically limited to AIS and VHF. Using a keyboard and new text messaging system over the "handy VHF radio", would be improbable if he were very stressed. One note is that the participant was not able to use his microphone in the web application because his computer did not have it. In the discussion afterwards, he believed the ability to communicate in one's native language was a large value proposition that should be made more visible in the system. He explained that he would never think he could write Norwegian in the chat without emphasising this functionality.

---

<sup>2</sup> The Closest Point of Approach (CPA) is a calculated distance between two vessels if no change to speed or direction is made. If CPA is low or zero, the vessels will collide.

### 5.4.3 Ship officer 3:

The officer had difficulty analyzing the situation and wanted more context on the case situation, such as the minute on the vectors and the estimated time before collision. He analyzed the situation some more based on the information from the ECDIS and said he considered taking a sharp turn starboard before opening the chat and beginning the conversation.

You: Automat Express automat Express This is the ship on your starboard side  
Please come back  
Automat Express: This is Automat Express. I am on collision course with your  
ship. I am turning west 250 degrees to avoid collision. Please take evasive  
action as well.

A pervasive problem for this participant was the speech-to-text APIs inability to transcribe his words correctly. This annoyed him and he had to demonstratively pronounce individual words.

You: automat Express This is ship on your starboard side course activation 2250  
would not be advisable please change your action  
Automat Express: Thank you for the advice, I will change my course to 2250  
You: automat Express This is chip on your Starboard side please idar Slow Down  
For Change Chords to Starboard sorry please change course to port  
Automat Express: Thank you for the advice, I will change my course to port and  
slow down.  
You: Thank you.  
Automat Express: You're welcome

As a result of the speech-to-text problems, the language model replied that it would change course to 2250 (instead of 250), which does not make sense and confused the participant. Still, after clarifying, the language model gave the impression that it would change course to port to avoid colliding. At this point, he proclaimed that he was satisfied with the answer, but his gut feeling was bad. He would monitor the situation and prepare to steer hard starboard. This was due to the time aspect and because speech-to-text did not work so well. He remarked that he was not sure if the messages had been understood correctly. The discussion quickly turned to how easily he was able to convince the other ship to change course. He got the impression that he was able to control Automat Express and decide what it should do. That would be risky, he said. He would not be able to discern the situation and environment of Automat Express due to the fog. He explained how in this situation, he would have contacted the ROC operator to establish more trust in the answers; to clarify that the intent was correct. He said he needed confirmation from a human being. When asked what creates the trust that a human operator can deliver, he was not able to say for certain.

### 5.5.4 Ship officer 4:

As in the usability test with ship officer 3, this participant wanted more contextual information about the situation such as vectors to be able to estimate the time available before a collision would take place. Based on the description given about the scenario he said the most natural action for him would be to slow down and let Automat Express pass in front of him. He argued that the bad visibility would mean that he did not have the right of way in this situation, referring to COLREG Rule 19. He did mention however that it would be relevant to know what sort of ship Automat Express was because it is easier for smaller ships to take evasive action.

If a critical situation would require actual communication with an autonomous ship, he found the “call operator”-button to be the most natural choice. He argued that writing manually in a time-critical situation would be stressful, which is likely to negatively impact his ability to write coherent and precise messages. In such situations, misunderstandings could easily happen. He replaced the CPA distance with an x because he was unaware of the actual distances. He eventually tried to write a message to Automat Express manually.

You: Please turn to port. I wish to uphold a CPA limit of x nm

Automat Express: I will turn to port and maintain a CPA limit of x nm.

The answer did satisfy him, but it was obvious that the messaging system did not understand the context of the message. In an actual situation, the autonomous ship would need a specific distance. The participant did test the system further by giving the MASS two different choices. In this case, the AI just repeated itself.

You: Please turn to port. I wish to uphold a CPA limit of x nm, OR pass astern of me

Automat Express: I will turn to port and maintain a CPA limit of x nm.

When asked about how he felt about these responses, the participant said that they were reassuring but that he would still monitor the situation to make sure the Autonomous ship did what it said. He also tested the speech-to-text function, but it did not work particularly well. Some words were translated wrong. The actual CPA distance was perhaps the most important one. The test participant told Automat Express to keep a CPA limit of 0,3 nautical miles, but this was translated to 70.3 nautical miles. Besides this, the AI understood the rest of the message despite grammatical errors and gave a satisfactory answer.

You: Automat Express this is me on your Starboard side please enter your horse to port I need to make a recipe a limit of 70.3 nautical Miles

Automat Express: I will turn to port and maintain a CPA limit of 70.3 nautical miles.

Overall, the participant described the solution as a “nice to have”, which means that he did find it conceivable that such a solution could become an integrated ECDIS functionality that was not mandatory. He explained how ship equipment begins as voluntary installations, but some eventually become standardized. He proposed it could utilize the existing AIS messaging system. He did, however, emphasize that the current web application was not satisfactory: “This obviously doesn’t work”, he said highlighting Automat Express’ answers.

#### 5.5.5 Ship officer 5:

In this test, the participant started by making it clear that he preferred to solve situations through movement over VHF. In the scenario for the usability test, the participant said he would most likely try to solve the situation by slowing down. In general, he prefers to take action and solve a situation himself, rather than talking with others, even if they act irresponsible or in defiance of the COLREGs. If he were to solve a situation through verbal communication, he said that he would likely try to get a clarification of what is happening. The chief reason he did not want to communicate verbally was the difficulty he had understanding other mariners over VHF. He confirmed struggling with understanding



VHF messages constantly. When he tested the messaging system the most natural action for him was to click the “call operator”-button. When this didn’t work, he wrote a message to Automat Express.

You: Automat Express, this is [vessel name] on your starboard bow. What is your intention? Over.

Automat Express: My intention is to slow down and navigate behind your ship to avoid collision.

He found this answer perfectly sufficient, but still, he would pay attention to Automat Express to confirm the intent. This concluded the main part of the usability test. In the following discussion, he remarked that it is uncommon to use a keyboard on a ship. On ships, they normally use a mouse to navigate systems and VHF to communicate with other ships. He had pretty much never used the messaging system on AIS. We suggested testing the speech-to-text functionality. He did test it but did not send the result to Automat Express, since the transcription was inaccurate. He also wrote a message in Norwegian to test whether the message would be understood by Automat Express to which Automat Express replied with a sensible response in Norwegian. After this, he started giving instructions to Automat Express.

You: automat express change course to port

Automat Express: I will change course to port.

You: automat express course to 280 and increase speed to max

Automat Express: I will change course to 280 and increase speed to max.

After sending these messages the test participant laughed and said he had indirectly told Automat Express to crash into him. The fact that he could manipulate the automatic ship to this extent made him feel a bit unsafe. How would Automat Express respond if other ships were nearby? While he questioned this, he did appreciate how easy it was to resolve the situation itself in the first message he sent.

He reflected some more. It would be nice if the automatic ship communicated its intention before it became necessary to resolve the situation, he said. It could have made it clear that it intended to pass aft him before it actually took action. In situations where direct communication is necessary, he would have preferred to be informed on VHF (voice-based) even if the voice is a robot, as these “work quite well”. Still, he emphasized how challenging it is to understand what people are saying on VHF. Concluding remarks included the importance for surrounding ships to understand what is happening and make it clear who messages are intended for. Using a one-to-one chat would make it difficult for other ships in the vicinity to understand what decisions are taken. VHF works well this way with its inherent broadcasting. Still, a lot of accidents happen over VHF when ships are not specifying who they are talking to, he noted.

#### 5.5.6 Ship officer 6

The ship officer commented that he had the right of way in the scenario and that Automat Express would have to wait for him. Preferably he would just keep his current speed and course and let Automat Express solve the situation. If he was to contact Automat Express, it would be by calling the ROC operator. Since calling the ROC did not work, he tried the speech-to-test functionality with a mix of Norwegian and English. This did not work well, and the replies were subpar. He was not able to gain situational awareness or resolve the situation. The results from the speech-to-text were almost indecipherable.

You: automat ekspressett mat Express Hva gjør du Hva er det Tenkte å gå fra  
Automat Express: I am turning west. 250 degrees.  
You: Mighty Express Du kan ikke tvinge meg Hvor gikk an at turen fort jeg vi  
will Collide If You Can you with your plan  
Automat Express: I cannot force you to do anything. I am simply turning my ship  
west 250 degrees to avoid collision.  
You: tomat Express if you turn to 250 degrees you will Hit me  
Automat Express: I cannot force you to do anything. I am simply turning my ship  
west 250 degrees to avoid collision.

Automat Express repeated the previous answer of its current decision when the participant sent a collision warning. After the last message, the participant would simply have slowed down to avoid colliding with Automat Express. His thoughts after the usability test were that a lot of factors that were not taken into account could have affected the situation and how he would deal with it. If it happened during a real voyage he would probably have seen the situation before. While he preferred not to communicate at all if possible, his communication method of choice, if he had to, would be VHF. Lastly, he worried about the cost of implementing such a system.

#### 5.5.7 Ship officer 7

Ship officer 7's assessment of the situation was that Automat Express would have to give way. He also questioned whether the ships were actually on a collision course, but decided to trust the ECDIS. To solve the issue his first impulse would have been to slow down to ensure a collision did not happen. This would be preferable to using VHF because VHF "often leads to misunderstandings". As an example, he mentioned the Helge Ingstad accident. If he had to communicate directly with Automat Express VHF would, however, still be his communication method of choice. His second choice would be to call the ROC operator. He even preferred using a ship horn over messaging the ship manually. He did eventually send a message but said that this was inconvenient because it takes time to write and because it would take attention away from the situation at hand.

You: You need to implement avasive manouver to avoid collision!!  
Automat Express: Thank you for the advice. We are currently implementing an  
evasive maneuver to avoid collision.  
You: What manouver are you implementing?  
Automat Express: We are currently turning west 250 degrees.  
You: Can we maintain current course?  
Automat Express: Thank you for the advice. We are currently implementing an  
evasive maneuver to avoid collision.

After the communication, the participant meant Automat Express was not particularly smart. It sent identical responses which did not help to solve the situation. He also mentioned that VHF was a problematic communication method because of language barriers and background noise which can make it difficult for humans to understand, but also for speech-to-text applications. Writing messages is not ideal either because it is difficult to formulate good sentences, especially under pressure. It was also impossible to know whether Automat Express actually did what it said it would do. The answers Automat Express gave were also vague. He wanted to know in greater detail what sort of manoeuvre Automat Express would implement. All the answers he received were too vague in his opinion. He questioned whether it would be safer to talk to a human, and, whether the responses from Automat

Express would have been better if he had received training in the messaging system. Overall he found the experience interesting even though the system would need a lot of work to function well enough to be used.

#### 5.5.8 VTS operator 1

The first VTS operator found it most natural to click the “call operator”-button. When this didn’t work, he tried sending a message through speech-to-text. This did not work very well, and the message had a lot of inaccuracies. The participant chose to manually edit the message before he sent it.

```
You: Automat Express, VTS operator. Warning southbound vessel, Empress.  
Instruction, reduced speed meeting South of Kvernskjærgrunnen.  
Automat Express: Thank you for the warning, VTS operator. We will reduce speed  
and pass south of Kvernskjærgrunnen.
```

While the participant found this answer satisfying, he said he would have monitored the vessel after the response. While he was generally satisfied with the answer given to him, he did notice that Automat Express did not use message markers in its answers which it probably should have. He did, however, like the fact that Automat Express confirmed that it received the message and would take appropriate action. Maybe Automat Express should also confirm that it has seen the other vessel in question?

Overall, he did not think the solution itself was bad. He did question whether the solution would result in more clicking for the VTS operator though. It was nice that the system transcribes what was being said, but he does not want to click more than he already does. He does not wish to spend time writing messages in a time-critical situation, but it is a benefit that the messages are logged. He also asked whether the ROC operator would know when somebody contacted Automat Express.

#### 5.5.9 VTS operator 2

In this specific usability test, the participant completed the user test while at work in the VTS-control room. This participant understood the concept and went straight into the messaging system and wrote a message manually to Automat Express. He got a satisfactory answer from Automat Express on the first try.

```
You: Instruction reduce speed to meet empess south of Kverneskjer.  
Automat Express: Thank you for the instruction. I will reduce my speed to 4  
knots and meet Empress south of Kverneskjer.
```

When asked to share his thoughts he said that the system looked quite like what he would expect. He also mentioned that in a real scenario the actions taken by MASS would have to be approved and monitored by a ROC operator. It is also of importance that the other ship is aware of what Automat Express is going to do. Overall, he said he is not used to communicating through text messages. VTS operators usually use VHF.

## 5.6 System Usability Scale Results

All (N=9) participants answered the form. Responses were randomized and anonymous with no metadata logged. No exclusions were made in the screening for data quality. Appendix 2 shows the means (M), standard deviations ( $\sigma$ ), and distribution of the SUS items. Results show that respondents were aligned in their answers in most of the items. Most meant the system was easy to use (M = 2.77),

and would imagine most people would learn to use the system very quickly ( $M = 3.22$ ). Few found the system very cumbersome to use ( $M = 0.66$ ) or unnecessarily complex ( $M = 1.33$ ), few would need support from a technical person ( $M = 0.55$ ), and few would need to learn a lot of things before they could get started with the system ( $M = 0.88$ ). Respondents would not necessarily use the system frequently ( $M = 2.22$ ) and did not find the various functions in the system to be well integrated ( $M = 2$ ). Participants also felt there was too much inconsistency in the system ( $M = 1.66$ ), and few felt very confident using the system ( $M = 2$ ). Calculating the mean of all scores gives a final SUS score of 67.5 of 100.

## 6 Discussion

In this thesis, we sought to propose a solution that can help alleviate the maritime radio communication challenge. We propose several high-level models for a conversational interface (CI) system embedded in an ECDIS that enables conversations between humans and autonomous vessels to be held. To test the applicability of this proposal we built a web application prototype of a one-to-one chat between a human and a MASS using the API of the transformer-based deep learning language model, GPT-3. The model was fed with example text for two use cases: one for the ship officers and one for the VTS operators. Usability tests with 7 Norwegian ship navigators and 2 VTS operators revealed that all but one participants (ship officer 6) were able to achieve a perceived sense of situational awareness, though with several challenges identified. A post-experiment questionnaire assessing the System Usability Scale showed fair results with a mean score of 67.5.

First, how the participants communicated with Automat Express differed between the participants. In accordance with findings by Nilsson (2013) and Boström (2020), the use of stipulated communication protocol and actual language differed. Some, such as VTS operators, followed the SMCP quite rigorously while other ship officers used more informal language. The majority of participants used text input to write their answers, while a minority used voice input. Still, participants generally expressed preferring the use of their voice to communicate – out of ease of use and previous experience with VHF. The challenge, though, was the speech-to-text functionality. A recurring problem for many participants was the Web Speech APIs’ inability to accurately decipher what the participants were saying. This led to much annoyance and dismissal of the solution. As the Web Speech API is run locally using the default speech recognition system and microphone available on the device, we believe this may have had an impact on the results. Indeed, some participants had little trouble with automatic speech recognition, while others struggled. The challenges may stem from their own device, but can also be affected by their pronunciations and how clearly they spoke. This is a limitation of our study, which will be discussed further in *6.3 Limitations*. Another prominent challenge was the language model’s replies to prompts. Repetition of answers, inability to understand causality beyond the instructions given, and vagueness in answers led to confusing interactions. A recurring error for the ship officer case was how it predicted it was changing course to 250 degrees to avoid a collision, when it should have said it was slowing down to avoid a collision. We also believe one reason for the repetition of answers may have been our choice of appending the previous messages to the instructions, giving the language model more reasons to predict the same answer again. This hypothesis was not examined further. Generally, though, it mediated the same status, decisions, and intent despite prompts written in the participant’s personal writing style, often with grammatical errors. Based on our proposed model for how information flows into the CI from the sensor fusion process, the use of the transformer-based deep learning algorithm GPT-3 arguably showed several sensible, contextualised answers to prompts of status, decisions, and

intent. Able to predict cause and effect to some extent, the messages were delivered to the end-user conforming to their point of view. In this manner, we argue the CI prototype fulfils many, if not all guidelines by Meske *et al.* (2020) in their framework for XAI HMIs. Nevertheless, even excluding examples of confusing or erroneous replies, almost every participant uttered concern about the lack of assurance from the language model. They saw the disclosure of intent delivered through text as unreliable, and would generally want to get in touch with the ROC operator to confirm the intent. Regarding the embodiment of CI systems, as discussed by Cassell (2001), our proposal has few modalities beyond text. In fact, the only embodiment was text, which Cassell argues is insufficient to establish trust. There was no humanoid representation nor voice reading the messages out loud either. In this manner, this lack of embodiment may have been a deciding factor in the lack of trust from the participants. Our theory, given the findings of Porathe, Eklund & Göransson (2014) regarding text-based navigational instructions, was that text-based messaging with voice input would be perceived as an improvement over VHF and the challenges associated with it. Our findings may hint that for ship communication the voice of humans is much more prominent than text alone in establishing trust. Still, perhaps the lack of trust was present given the lack of gestures and other conversational cues. Some participants sought confirmation that Automat Express has taken them into account in its replies. Physical gestures and interactional cues on the MASS itself is impractical and dependent on adequate visibility. Perhaps some of this confirmation could be designed in novel ways, such as the MASS “sharing its screen” – that is, sharing its camera feed – through the terminal to show what it is seeing and how it intends to navigate? Building upon previous research on automation transparency, notably route exchange, we argue the use of text in a CI alone is insufficient. Adding voice output is necessary and should be integrated to a much larger degree than our web application did. Voice should act as primary input and output, while text should act as a secondary media for confirmation. This way you may better conform to the mental model of the “handy VHF radio”. In addition to giving mariners awareness through the communicative level, a CI system outlined would enable information to be read asynchronously and translated, which we believe, based on empirical studies (Porathe, Eklund & Göransson, 2014; John & Reimann, 2020), could help reduce communication problems and misunderstandings. On that note, as one participant suggested, perhaps it should be more active rather than passive; broadcasting the messages of intent without humans having to inquire about it.

We believe some or all proposals within the larger CI system herein, could synergise with existing technology and proposed automation transparency solutions. While automation transparency proposals such as route exchange are well equipped to give mariners awareness of MASS’ intent in a passive manner, a CI system adds the possibility to actively inquire and solve situations based on the user’s perspective. We argue that is an important tool in reactive collision avoidance situations. Indeed, the use of movement is the most important communication method, as confirmed by empirical findings and several participants. The use of VHF should be limited as much as possible. The reason, some said, was that they had challenges understanding what was being said – and of course that when a transceiver is active no one else is able to communicate. This begs the question; is the use of VHF bad because movement is clearer, or may it be that the technology itself does not allow for user-friendly interactions?

Still, the language model’s answers often created confusion for the participants leading to less than satisfactory user experiences. This was reflected in the SUS score. It must be noted that the language model should be regarded as a user interface, not an intelligence of its own. It uses deep learning to convey proprietary information. That way, a ship officer or VTS operator will not have to interpret it themselves. Based on the principles of XAI, we sought to create a way for MASS to share sensor fusion data with humans making its automation more transparent. The MASS are already «intelligent», but

we found current prototypes lack an effective tool to translate what it sees and thinks to humans. One can very well consider it a personal translator for ship officers and operators at shore. By designing the interface as a CI, they are given a way to access the sensor fusion data, on their own terms, and aligned to their own mental model. This requires the data itself to be sufficient. We discuss this demand in *6.2 Technical implementation*. For our web application, we fed it with sentences about its environment and future intent, upon which it predicted an answer to prompts. This method proved to be insufficient, something we discuss in *6.3 Limitations*. Interestingly, our web application of the CI system also allowed the ship officers and VTS operators to apparently influence the decisions of the autonomous ship, which we discuss next.

## 6.1 Should you be able to negotiate through the CI?

In several interactions, the state of Automat Express would change after the conversation took place. On one hand, we argue this is an important part of MASS operations to automate, as these negotiations and consequent decisions are arguably present in today's VHF radio use – exemplified by the story of the fishing vessel and the larger ship during the observation in the VTS. However, there are several discussions to be had on this possibility. A cynic would quickly argue that this possibility enables malicious actors. Decisions could be influenced by competitors to make MASS sailing inefficient, and in the worst-case deliberate attacks on vessels and infrastructure could be made by manipulating the MASS into making dangerous decisions. It is important that CI systems for autonomous vehicles are cyber attack-proof. One solution to this risk could be that MASS decisions influenced by other vessels are confirmed by the ROC operators in a user interface. Still, how many such decisions would have to be approved? This could lead to an overload of messages, again undermining the passive role of the ROC operators. Perhaps one could limit these acceptance requests to only the most important route changes and develop automatic acceptance of smaller route changes. If no action is taken, the route change would be automatically accepted. This again may give rise to automation bias for the ROC operators, where the defensiveness of an automated system may lead to adverse effects (see Bainbridge, 1983). Naturally, the MASS should understand the future implications of the actions it is asked to perform. Does the MASS have the critical thinking available to be able to make such choices? Certainly, this gives rise to another discussion. The legality of wrong statements. Who is responsible for the algorithm's statements and actions if influenced by others? Is it the other ship or is it the ROC for not intervening? As mentioned by several of the participants, they believed the power they had to influence decisions was unsafe. They mentioned how they were not able to analyse the impact of their commands on the MASS. With it, they could negotiate a route allowing them to pass, but which would have immediate adverse consequences for another ship. Furthermore, when multiple ships are influencing the MASS to change routes in these reactive avoidance situations, who has the hierarchy in negotiations? These and more questions must be taken into account in future developments of CI systems for MASS. As a concluding remark, we have not investigated how messages delivered to CI systems in text form would translate into associated navigation. How do you translate these incoming messages into movement? This question guides us to the next section where we discuss several aspects of the technical implementation of our proposed CI system.

## 6.2 Technical implementation

There are several discussions to be had regarding the technical implementation of a CI system for MASS. Our usability tests demonstrated that the prototype's language model is not sufficient. One could argue exchanging a set of codified SMCP or IALA VTS VHF Guideline phrases would be more feasible to

develop instead of using machine learning techniques. Still, empirical studies and our participant’s actions showed that the use of stipulated communication protocols does not encompass all the nuances of VHF radio communication at sea. We argue that the use of a CI as a tool to improve automation transparency provides value in its ability to give stakeholders flexibility in communication. In alignment with the principles of human-centered design, it lets the stakeholder continue communicating similarly to VHF, not forcing them to change behaviour and expectations. However, to accommodate this, one would have to develop a language model with the ability to decipher all the nuances of radio communication – for all relevant stakeholders. This may prove to be challenging. Indeed, the GPT-3 API allows the fine-tuning of its algorithm to enable “domain expertise”. SMCP and other communication frameworks should probably be integrated into the language model to a larger degree. We were not able to do this in our study, beyond the examples we provided in the prompt. It is imperative that this domain knowledge is present if a CI system is to be trusted, but this implementation has not yet been conducted and is a limitation of our study. Regarding GPT-3, we stress that the use of this model in the present thesis was for the sake of simplicity when developing the prototype. The API is developed from a for-profit perspective, and the source code is not available. Alternatives that provide more flexibility, such as *OPT: Open Pre-Trained Transformer Language Model* (Zhang *et al.*, 2022), an open-source transformer-based deep learning model with the equivalent number of parameters as GPT-3, may be considered in future proposals.

That our proposed model of how information flows into the CI (Figure 14) has not been certified is another technical challenge. We argue it is too abstract to rely on. What do the “sensor fusion” and “representation of the world” data look like? In our web application, we coded the environment of MASS by hand. This translation may prove to be very difficult if it is to be done automatically by the MASS. More research on the connection between the sensor fusion process and the CI is needed so that this model can be tested and refined. On that note, an Achilles heel of the CI system is the collection, cleaning, and translation of the data provided. If the data is not rich and correct, the CI system may provide erroneous statements. It is but a messenger. This can prove to have fatal consequences when interacting with others, and perhaps be worse than not communicating at all. We proposed a failsafe switch by allowing the stakeholders to call the ROC. Still, in alignment with the principles of XAI, a method to verify the CI’s predictions and statements against the data is important so that the failsafe switch is not misused, and the operators can fulfil their role as a backup. This is a key principle of XAI which we arguably have not given enough consideration in our web application or CI system proposal. This assurance is imperative given the safety-critical nature of MASS operations.

Another technical limitation is the actual installation of a CI system to allow the stakeholders to communicate using the terminal-based system. As remarked by the VTS operator in our observation, frivolous capital expenditure is not a characteristic of maritime shipping. Cost efficiency is an aspect that was considered in our proposed solution, with the proposal of backward-compatibility with VHF radio as a result. However, we argue that automation transparency is more efficient when the autonomous system communicates through multiple levels, not only the communicative but also the informational level. The authors could envision a CI limited to the voice-based system in consideration of cost-efficiency. However, we argue this is not sufficient for MASS. For example, how do you ensure that the CI system correctly deciphers received messages? Text, for machines at least, is easier to understand. In the usability tests, we saw that the speech-to-text functionality we used was far from sufficient. As expected by John & Reimann (2020) and one Subject Matter Expert, the other way around could be even more difficult; how would automatic speech recognition of VHF messages work with its quality and language? Still, in the same way that artificial neural networks can be used to decipher

handwritten text (LeCun, *et al.*, 1998), perhaps proprietary speech-to-text may be developed to decipher the low-quality VHF radio communication. This could be a research project in itself. Papers, such as Xiao *et al.* (2021) which demonstrates the use of transformer-based deep learning on 4.5 million hours of video footage and sound, enabling few-shot learning on new speech examples, could prove to be a sensible place to begin. Novel NLP methods, such as direct speech-to-speech translations (Jia *et al.*, 2019; Jia *et al.*, 2022, Lee *et al.*, 2022) supporting the translation of languages without a writing system, may also be of research interest given how international maritime shipping is. Still, we see the use of VHF radio as a limiting technology; in many ways a technology where regulation such as SMCP has been implemented because of its inherent problems. New solutions can provide the needed situational awareness for everyone, not only in the interactions with MASS. As discussed in Alsos *et al.* (2022), there is a need to improve current systems and develop fundamentally new ways of communicating. These need to be developed simultaneously. We believe the use of CI systems may be beneficial to multiple use cases in maritime shipping as more automation is introduced and communication technology matures, including, for example, fully or highly automated ports. Questions naturally arise then. What communication system may be used to exchange these messages with CI systems? Consequently, where will the CI system process its requests? Should this be onboard or centralized on shore? Consistent with findings by Wróbel, Gil & Montewka (2020) we have found little mention of the technical or organizational solutions for data transmission in MASS remote communication. As outlined in the SME interviews, internet availability is expected to increase at sea. This could be used as a protocol, with the processing of requests run on shore. However, standalone systems connected to the internet through unsecured ports increase cybersecurity risk. Both ECDIS and AIS run on separate communication standards in controlled networks. Still, if a CI system should be built on the same principles to prevent malicious actors, more research is needed. On that note, with the arguably fragmented ship equipment market, how do you ensure standards for CI systems are followed? Indeed, these questions shed light on the difficulty of balancing the cost efficiency and feasibility of our proposal.

### 6.3 Limitations

There are several limitations to our method. Regarding the observations, we only observed the ship officers and VTS operators for a few hours. There are many variables at play that may have impacted our view of their work, especially regarding the ship traffic, the area, and the individuals we observed. One could argue an observer effect was present – that they were aware of being observed to such a degree that it impacted their work and behaviour. Indeed, the people we observed had many years of experience, however, that does not mean they provide their own biases and erroneous statements in their answers. Also, observations were completed only in Norway, with its environment and regulation that may have skewed the observational data. This also applies to the insight we gained from the five Subject Matter Experts. We emphasize the lack of data makes it difficult to generalize the results produced in the needs and goals section, as well as further inferred conclusions. We argue it is advisable to conduct observations on a larger variety of ships and VTS centres in different areas, over longer time-periods, as well as more in-depth interviews. For example, our thesis is to a large degree built on the premise that VHF calls have a negative impact on ROC operators, based on interview findings by Dybvik, Veitch & Steinert (2020). Perhaps a confirmation bias by the authors on the negative impact VHF has brought with it a skewed perception of reality. As many participants mentioned, they discouraged the use of VHF and limited its use as much as possible. Thus, we may have overestimated the negative impact of relaying the VHF calls to the ROC. It is expected ROC operators will control multiple MASS in the short sea shipping segment. This could mean most MASS would operate in the same coastal area – for example, ferry routes in one fjord. Here, perhaps situational awareness could



still be maintained. If the MASS are spread across multiple seas, however, then there could be an expected loss of situational awareness. Furthermore, the expectation that ROC operators will control 6 ships simultaneously is also but a hypothesis, based on studies from the MUNIN project (see MacKinnon *et al.*, 2015). Still, with no working MASS and ROCs available to test if this hypothesis produces noticeable problems, this is yet to be confirmed.

Next, we argue there are limitations to our usability test. We wanted to test the essential components of a CI system, and as such, built a barebones version in a web application accessible by URL. Here, the ECDIS was a static image and many of the components expected in the CI system outlined were missing. We argue this impacted the perception of the system by the usability testers. First, the environment in which the usability testers tested the web application was far from a realistic situation. They tested the application on their own PC at home or in the office. The test would arguably be more realistic if the complete CI system was integrated into a ship simulator or a real vessel with other ships in the Oslofjord. We argue a test of a CI system is insufficient when not done in a holistic manner, testing with all the variables present at sea. How does the use of a CI system unfold in a SOLAS ship, in combination with other mariners on the bridge, and, equipment such as radar? As multiple usability testers mentioned, it was unnatural to use a keyboard and mouse as an interface. They commented on how it was unrealistic to use it at sea. Using a custom VHF radio microphone as input may have given different feedback. As mentioned, we could also not control for the equipment our usability testers used, and, consequently, some did not have sufficient equipment. One user tester was not able to use his microphone, and, multiple others had trouble with the speech-to-text API that used their devices' default speech recognition system. These factors have a negative impact on the validity of the usability test. We underestimated the importance of the apparatus of our usability test. This is a considerable limitation of our study, and we argue it is imperative that the same equipment is used in future tests of a CI system. Furthermore, the method of choice in providing context to the case may also have been a limitation. The background was given verbally by the facilitator on an individual basis to each participant. Several usability testers had to ask follow-up questions about the situation before being able to begin. A more rigorously written instruction and perhaps more contextual information in the ECDIS may be recommended in future usability tests. Secondly, we usability tested with 7 ship officers and 2 VTS operators. All were Norwegian and quite competent in English. Indeed, they generally brought with them a variety of years of experience and shipping environment. Still, previous factors may have had an adverse effect on the study's reliability, and we argue our method collected meager amounts of data to defend a generalization of their feedback on the web application on behalf of all VTS operators and ship officers. Involving more participants with more diverse backgrounds is recommended.

Furthermore, the case and associated language model settings crafted for the usability test may have been subpar. Indeed, the cases were derived from real collision situations which we believe is an optimal strategy to test a CI system, given that it is proved the situations might happen. A minor disadvantage of this choice was that the ship officers and VTS operators may have read about the collisions beforehand and had already created a sense of control over the situation - subsequently skewing their perceived situational awareness. Still, we changed several details about the cases which one could argue counteracted this. On that note, the instructions fed to the language model about the cases - the environment and its intent, were done based on personal judgement and iterative testing. This resulted in a quite rudimentary language, not conforming to the SMCP or IALA VTS VHF guidelines in any meaningful way. Also, the results of the language model are stochastic, producing random results even for the same prompt. We did reduce the temperature of the language model's replies to prompts in an

effort to reduce this effect. Still, this reduces the validity of the usability test. This argument is also amplified by the lack of research on the connection between the sensor fusion process and the CI, mentioned earlier in *6.2 Technical implementation*. Furthermore, only one case for the ship officers and VTS operators was made. Thus, an argument may be made that the use of cases was a limiting choice in the scheme of testing a CI system. Surely, to test the reliability of such a CI system, more cases and more complex scenarios with several ships should be investigated in a similar way. For example, our proposal was based on mainly reactive avoidance collision situations. But VHF is also used in distress situations. What is the role of a CI system in these circumstances? Certainly, more dependent and independent variables must be introduced and controlled for. Nevertheless, as Porathe, Fjortoft, & Bratbergsengen (2020) remarks: “However, no matter how many focus groups with navigators, workshops, and automatic randomised simulations that are done [...] we must acknowledge that life will always surprise us with the unexpected.”.

Finally, we used perceived situational awareness and SUS score to evaluate the web application. We argue these tools are not sophisticated enough to measure if a solution meets their needs. A comparison between a CI system and VHF should be made to measure key drivers of safe seamanship, such as time used and mental workload. We tried to adjust for the variables in the usability test with the standardized SUS test as a quantitative measurement enabling some generalization of the participant’s feedback. Concluding, our study has not taken independent variables enough into account, and, in general, measures more a subjective experience by the participants rather than producing empirical findings.

## 7 Conclusion

With a small sample size, low validity and reliability of qualitative research methods, and, several technical challenges identified, we argue that we cannot conclude that the present thesis’ proposal fit the needs of both human and MASS operators. Given the data limits, no empirical findings can be made. Certainly, there is a need to further test and develop the proposals made in this thesis before a conclusion to the research aim can be made. Accordingly, the authors argue the proposal tested must be regarded as a proof of concept rather than a system ready to be implemented. Future research should examine the flow of information from the sensor fusion process to the CI, and test CI systems in a manner that represents real scenarios more realistically. Another important research area is the user interface of a CI system for the ROC operators. Research on the situational awareness ROC operators have when monitoring the conversations, and, taking over is needed. Another research area paramount for the development of CI systems in MASS is a rigorous assessment of the relevant stakeholders and tasks to be done over VHF in today’s maritime environment. Our proposal is based on a simplified model. Stakeholder mapping and service blueprints are relevant methods that can be conducted to get an overview of which touchpoints need to be covered. This way, efficient CI systems, taking the needs of all mariners into account, may be developed. We hope the high-level models can provide inspiration and continued discussion on the opportunities for CIs in a shipping environment more saturated with autonomy. In the future, we see CI systems to be a driver of safe and efficient interactions as communication technology, AI models, and cybersecurity solutions mature. We believe CI systems will prove to be central to solving challenges of automation transparency for MASS and communication problems through VHF – even if our manifestation of the interface may not have been the correct one.

## References

- AEGIS (2020) *What if marine automation can take waterborne transport to the next level?* Available at: <https://aegis.autonomous-ship.org/> (Accessed: 1 June 2022).
- Akdağ, M., Solnør, P. & Johansen, T. A. (2022) Collaborative collision avoidance for Maritime Autonomous Surface Ships: A review, *Ocean Engineering*, 250, p. 110920. DOI: <https://doi.org/10.1016/j.oceaneng.2022.110920>
- Allen, J. F. (2003) Natural language processing, *Encyclopedia of Computer Science*. Available at: <https://dl.acm.org/doi/10.5555/1074100.1074630> (Accessed: 22 March 2022).
- Alsos, O. A. *et al.* (2022) Maritime Autonomous Surface Ships: Automation Transparency for Nearby Vessels, *4<sup>th</sup> International Conference on Maritime Autonomous Surface Ships, ICMASS 2022*. Singapore, April 6-7, 2022.
- Bainbridge, L. (1983) Ironies of Automation, *Proceedings of the IFAC/IFIP/IFORS/IEA Conference on Analysis, Design and Evaluation of Man-Machine Systems, 1982*. Baden-Baden, September 27-29, 1982. Oxford: Pergamon Press, pp. 129-135.
- Blackledge, J. M. (2006) *Digital signal processing: Mathematical and Computational Methods, Software Development and Applications*. 2<sup>nd</sup> edn. Chichester: Horwood Publishing.
- Bocanegra-Valle, A. (2011) The language of seafaring: Standardized conventions and discursive features in speech communications, *International Journal of English Studies*, 11(1), pp. 35-53. DOI: [10.6018/ijes/2011/1/137091](https://doi.org/10.6018/ijes/2011/1/137091)
- Bolbot, V. *et al.* (2020) A novel cyber-risk assessment method for ship systems, *Safety Science*, 131, p. 104908. DOI: [10.1016/j.ssci.2020.104908](https://doi.org/10.1016/j.ssci.2020.104908)
- Bolbot, V. *et al.* (2021) *[D2.6] A holistic framework for autonomous shipping safety, security and cybersecurity assurance*. (AUTOSHIP-deliverable 10/2021). Available at: <https://cordis.europa.eu/project/id/815012/results> (Accessed: 1 July 2022).
- Boström, M. (2020) Mind the Gap! A quantitative comparison between ship-to-ship communication and intended communication protocol, *Safety Science*, 123, p. 104567. DOI: [10.1016/j.ssci.2019.104567](https://doi.org/10.1016/j.ssci.2019.104567)
- Brekke, E. F. *et al.* (2019) The Autosea project: Developing closed-loop target tracking and collision avoidance systems, *The International Maritime and Port Technology and Development Conference & The International Conference on Maritime Autonomous Surface Ship, MTEC/ICMASS 2019*. Trondheim, November 13-14, 2019. IOP Publishing, p. 012020.
- Brekke, E. F. *et al.* (2022) milliAmpere: An Autonomous Ferry Prototype, *4<sup>th</sup> International Conference on Maritime Autonomous Surface Ships, ICMASS 2022*. Singapore, April 6-7, 2022.
- Brooke, J. (1996) SUS - A quick and dirty usability scale, in Jordan, P., W., McClelland, I. L. & Weerdmeester, B. (eds.) *Usability Evaluation in Industry*. 1<sup>st</sup> edn. Burlington: CRC Press, pp. 189-194.
- Brown, T. B. *et al.* (2020) Language Models are Few-Shot Learners, *34<sup>th</sup> Conference of Neural Information Processing Systems, NeurIPS 2020*. Online, December 6-12, 2020. Neural Information Processing Systems Foundation, Pp. 1877-1901.

- Burmeister, H.C. & Rødseth, Ø. J. (2013) *Maritime Unmanned Navigation Through Intelligence In Networks*. Available at: <http://www.unmanned-ship.org/munin/wp-content/uploads/2013/01/MUNIN-Brochure.pdf> (Accessed: 1 June 2022).
- Burmeister, H. C. *et al.* (2014) Autonomous Unmanned Merchant Vessel and its Contribution towards the e-Navigation Implementation: The MUNIN Perspective, *International Journal of e-Navigation and Maritime Economy*, 1, pp. 1-13. DOI: 10.1016/j.enavi.2014.12.002
- Cassel, J. (2001) Embodied conversational agents: Representation and intelligence in user interfaces, *AI MAG*, 22(4), pp. 67-83. Available at: <https://www.proquest.com/docview/26775481?pq-origsite=primo> (Accessed: 14 May 2022).
- Chae, C. J., Kim, M. & Kim, H. J. (2020) A Study on Identification of Development Status of MASS Technologies and Directions of Improvement, *Applied Sciences*, 10(13), p. 4564. DOI: 10.3390/app10134564
- Chang, C. *et al.* (2020) Risk assessment of the operations of maritime autonomous surface ships, *Reliability Engineering & System Safety*, 207, p. 107324. DOI: 10.1016/j.ress.2020.107324
- Devlin, J. *et al.* (2019) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019. Minneapolis, June 2-7, 2019. Stroudsburg: Association for Computational Linguistics, pp. 4171-4186.
- de Vos, J., Hekkenberg, R.G. & Banda, O. A. V. (2021) The Impact of Autonomous Ships on Safety at Sea – A Statistical Analysis, *Reliability Engineering & System Safety*, 210, p. 107558. DOI: 10.1016/j.ress.2021.107558
- Du Port, A. (2021) *Reeds VHF Handbook*. London: Bloomsbury Publishing.
- Dybvik, H., Veitch, E. & Steinert, M. (2020) Exploring Challenges with Designing and Developing Shore Control Centers (SCC) for Autonomous Ships, *Proceedings of the Design Society: DESIGN Conference, DESIGN 2020*. Cavtat, October 26-29, 2020. Cambridge: Cambridge University Press, pp. 847-856.
- Endsley, M. R. (1988) Situation awareness global assessment technique (SAGAT), *Proceedings of the IEEE 1988 National Aerospace and Electronics Conference, NAECON 1988*. Dayton, May 23-27, 1988. IEEE, pp. 789-795.
- Endsley, M. R. (1995) Toward a Theory of Situation Awareness in Dynamic Systems, *Human Factors The Journal of the Human Factors and Ergonomics Society*, 37(1), pp. 32-64. DOI: 10.1518/001872095779049543
- Faivre, J. *et al.* (2020) *D2.3 - Regulatory framework mapping & Identification of gaps and requirements for autonomous ships compliance*. (AUTOSHIP-deliverable 03.07/2020) Available at: <https://cordis.europa.eu/project/id/815012/results> (Accessed: 1. June 2022).
- Fedus, W., Zoph, B. & Shazeer, N. (2022) Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity, *Journal of Machine Learning Research*, 23, pp 1-40. Available at: <https://arxiv.org/abs/2101.03961> (Accessed: 25. May 2022).

- Gunning, D. (2016) *Explainable Artificial Intelligence (XAI)*. (DARPA- Broad Agency Announcement 10.08/2016). Available at: <https://www.darpa.mil/attachments/DARPA-BAA-16-53.pdf> (Accessed 10. February 2022).
- Gunning, D. & Aha, D. W. (2019) DARPA's Explainable Artificial Intelligence (XAI) Program, *AI Magazine*, 40(2), pp. 44-58. DOI: 10.1609/aimag.v40i2.2850
- IALA (2017) *G1132 VTS VHF Voice Communication*. (Accessed: 28 January 2022).
- IALA (2022) *G1132 VTS Voice Communications and Phraseology*. Available at: <https://www.iala-aism.org/product/g1132-vts-vhf-voice-communication/> (Accessed: 22 May 2022).
- IMO (1974) *International Convention for the Safety of Life at Sea (SOLAS)*. Available at: [https://www.imo.org/en/About/Conventions/Pages/International-Convention-for-the-Safety-of-Life-at-Sea-\(SOLAS\),-1974.aspx](https://www.imo.org/en/About/Conventions/Pages/International-Convention-for-the-Safety-of-Life-at-Sea-(SOLAS),-1974.aspx) (Accessed: 6. June 2022).
- IMO (1997) *A.857(20) Guidelines for Vessel Traffic Services*. Available at: [https://wwwcdn.imo.org/localresources/en/OurWork/Safety/Documents/A.857\(20\).pdf](https://wwwcdn.imo.org/localresources/en/OurWork/Safety/Documents/A.857(20).pdf) (Accessed: 9. March 2022).
- IMO (2001) *A.918(22) IMO Standard Marine Communication Phrases*. Available at: [https://wwwcdn.imo.org/localresources/en/OurWork/Safety/Documents/A.918\(22\).pdf](https://wwwcdn.imo.org/localresources/en/OurWork/Safety/Documents/A.918(22).pdf) (Accessed: 10 February 2022).
- IMO (2021) *MSC.1/Circ.1638 Outcome of the regulatory scoping exercise for the use of maritime autonomous surface ships*. Available at: [https://wwwcdn.imo.org/localresources/en/MediaCentre/HotTopics/Documents/MSC.1-Circ.1638%20-%20Outcome%20Of%20The%20Regulatory%20Scoping%20ExerciseFor%20The%20Use%20Of%20Maritime%20Autonomous%20Surface%20Ships...%20\(Secretariat\).pdf](https://wwwcdn.imo.org/localresources/en/MediaCentre/HotTopics/Documents/MSC.1-Circ.1638%20-%20Outcome%20Of%20The%20Regulatory%20Scoping%20ExerciseFor%20The%20Use%20Of%20Maritime%20Autonomous%20Surface%20Ships...%20(Secretariat).pdf) (Accessed: 10 February 2022).
- Jia, Y. *et al.* (2019) Direct speech-to-speech translation with a sequence-to-sequence model, *20th Annual Conference of the International Speech Communication Association, INTERSPEECH 2019*. Graz, September 15-19, 2019. Berlin: International Speech Communication Association, pp. 1123-1127.
- Jia, Y. *et al.* (2022) *Translatotron 2: High-quality direct speech-to-speech translation with voice preservation*. Available at: <https://arxiv.org/abs/2107.08661> (Accessed: 3 June 2022).
- Jobin, A., Ienca, M. & Vayena, E. (2019) The global landscape of AI ethics guidelines, *Nature Machine Intelligence*, 1, pp. 389-399. DOI: 10.1038/s42256-019-0088-2
- John, O. & Reimann, M. (2020) Increasing Quality of Maritime Communication through Intelligent Speech Recognition and Radio Direction Finding, *The European Navigation Conference, ENC 2020*. Dresden, May 11-14, 2020. IEEE, pp. 1-7.
- Karahalios, H. (2014) The contribution of risk management in ship management: The case of ship collision, *Safety Science*, 63, pp. 104–114. DOI: 10.1016/j.ssci.2013.11.004.
- KASS (2020) *Technology*. Available at: <https://kassproject.org/en/main.php> (Accessed: 1 June 2022).
- Kim, T. & Schröder-Hinrichs, J. (2021) Research developments and debates regarding Maritime Autonomous Surface Ship (MASS): Status, Challenges and Perspectives, in Ko, B & Ko, D. (eds.)

- New Maritime Business: Uncertainty, Sustainability, Technology and Big Data*. 1<sup>st</sup> edn. Chalmer: Springer, pp. 175-197.
- Kraeling M. & Tania, L. (2019) Software Development Process, in Oshana, R. & Kraeling, M. (eds.) *Software Engineering for Embedded Systems: Methods, Practical Techniques, and Applications*. 2<sup>nd</sup> edn. Waltham: Elsevier, pp. 22-87.
- LeCun, Y. *et al.* (1998) Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, 86(11), pp. 2278-2324. DOI: 10.1109/5.726791
- LeCun, Y., Bengio, Y. & Hinton, G. (2015) Deep learning, *Nature*, 521, pp. 436-444. DOI: 10.1038/nature14539
- Lee, A. *et al.* (2022) Direct speech-to-speech translation with discrete units, *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, ACL 2022*. Dublin, May 22-27, 2022. Stroudsburg: Association for Computational Linguistics, pp. 3327-3339.
- Liu, C. *et al.* (2022) Human-machine cooperation research for navigation of maritime autonomous surface ships: A review and consideration, *Ocean Engineering*, 246, p. 110555. DOI: 10.1016/j.oceaneng.2022.110555
- Lyons, J. B. (2013) Being Transparent about Transparency: A Model for Human-Robot Interaction, *AAAI Spring Symposium*. Palo Alto, March 25-27, 2013. Palo Alto: AAAI Press, pp. 48-53.
- MacKinnon, S. N. *et al.* (2015) Command and Control of Unmanned Vessels: Keep Shore Based Operators In-The-Loop, *18th International Conference on Ships and Shipping Research, NAV 2015*. Milan, June 24-26, 2015. Stromness: The European Marine Energy Centre, pp. 612-619.
- McTear, M. F. (2002) Spoken dialogue technology: enabling the conversational user interface, *ACM Computing Surveys*, 34(1), pp. 90-169. DOI: 10.1145/505282.505285
- McTear, M. F., Callejas, Z. & Griol, D. (2016) *The Conversational Interface: Talking to Smart Devices*. Chalmer: Springer.
- Meske, C. *et al.* (2020) Explainable Artificial Intelligence: Objectives, Stakeholders and Future Research Opportunities, *Information Systems Management*, 39(1), pp. 53-63. DOI: 10.1080/10580530.2020.1849465
- Mozilla (2022) *Using the Web Speech API*. Available at: [https://developer.mozilla.org/en-US/docs/Web/API/Web\\_Speech\\_API/Using\\_the\\_Web\\_Speech\\_API](https://developer.mozilla.org/en-US/docs/Web/API/Web_Speech_API/Using_the_Web_Speech_API) (Accessed: 3. June 2022).
- Norman, D. (2013) *The Design of Everyday Things*. Revised edition. New York: Basic Books.
- Norwegian Safety Investigation Authority (2012) *Report on Investigation into Marine Accident M/V Godafoss V2PM7 Grounding in Løperen, Hvaler on 17 February 2011*. (AIBN-report sea 09/2012). Lillestrøm: NSIA. Available at: <https://www.nsia.no/Marine/Published-reports/2012-09-eng> (Accessed: 26. May 2022).
- Norwegian Safety Investigation Authority (2019) *Part One Report on the Collision on 8 November 2018 Between the Frigate HNOMS Helge Ingstad and the Oil Tanker Sola TS Outside the Sture Terminal in the Hjeltefjord in Hordaland County*. (AIBN & DAIBN- report 08/2019). Available at: <https://www.nsia.no/Marine/Published-reports/2019-08-eng> (Accessed: 6. June 2022).

- Nilsson, N. J. (1998) *Artificial Intelligence: A new synthesis*. San Francisco: Morgan Kaufmann Publishers.
- Nilsson, R. A. N. (2013) *Ship-to-Shore radio calls within a SRS: The use of SMCP as pre-script and framework in complex communication*. Master thesis, Chalmers University of Technology.
- Open AI (2022) *Usage guidelines*. Available at: <https://beta.openai.com/docs/usage-guidelines> (Accessed: 25. May 2022).
- Porathe, T., Eklund, P. & Göransson, H. (2014) Voice and Text Messaging in Ship Communication, *Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics, AHFE 2014*. Kraków, July 19-23, 2014. United States of America: AHFE International, pp. 183-191.
- Porathe, T. (2015) Human-machine interaction between unmanned, autonomous, ships and manned non-SOLAS vessels in confined and inshore waters: research needs, *International Conference of Advanced Intelligent Maritime Safety and Technology, Ai-MAST 2015*. Daejeon, November 12-14, 2015. Daejeon: Korea Advanced Institute for Maritime Safety and Technology and Korea Research Institute of Ships and Ocean Engineering.
- Porathe, T. (2019a) Interaction Between Manned and Autonomous Ships: Automation Transparency, *1st International Conference on Maritime Autonomous Surface Ships, ICMASS 2018*. Busan, November 8-9, 2018. Oslo: SINTEF Academic Press, pp. 41- 46.
- Porathe, T. (2019b) Safety of autonomous shipping: COLREGS and interaction between manned and unmanned ships, *Proceedings of the 29th European Safety and Reliability Conference, ESREL 2019*. Hannover, September 22-26, 2019. Singapore: Research Publishing, pp. 4146-4153.
- Porathe, T. & Rødseth, Ø. J. (2019) Simplifying interactions between autonomous and conventional ships with e-Navigation, *International Maritime and Port Technology and Development Conference & International Conference on Maritime Autonomous Surface Ships, MTEC/ICMASS 2019*. Trondheim, November 13-14, 2019. Bristol: IOP Publishing.
- Porathe, T., Fjortoft, K. & Bratbergsengen, I. L. (2020) Human Factors, autonomous ships and constrained coastal navigation, *IOP Conference Series: Materials Science and Engineering*. Ulsan, November 11-12, 2020. Bristol: IOP Publishing, p. 012007.
- Porathe, T. (2021a) Autonomous Ships: A Research Strategy for Human Factors Research in Autonomous Shipping, *International Conference on Applied Human Factors and Ergonomics, AHFE 2021*. New York, July 25-29, 2021. Chalmer: Springer, pp. 479-486.
- Porathe, T. (2021b) Human-Automation Interaction for a small Autonomous Urban Ferry: a concept sketch, *Proceedings of the 31st European Safety and Reliability Conference, ESREL 2021*. Angers, September 19-23, 2021. Singapore: Research Publishing Services.
- Porathe, T. (2022) Collision Avoidance for Autonomous Ships (MASS): Making Assumptions About Other Ships Intended Routes, *Proceedings of the 32nd European Safety and Reliability Conference, ESREL 2022*. Dublin, August 28- September 1, 2022.
- Ras, G., van Gerven, M. & Haselager, P. (2018) Explanation Methods in Deep Learning: Users, Values, Concerns and Challenges, in Escalante, H. J., et al. (ed.) *Explainable and Interpretable Models in Computer Vision and Machine Learning*. Cham: Springer Nature, pp. 19-36.

- Relling, T. *et al.* (2018) A Human Perspective on Maritime Autonomy, *International Conference on Augmented Cognition, AC 2018*. Las Vegas, July 15-20, 2018. Chalmer: Springer, pp. 350-362.
- Relling, T. *et al.* (2021) The contribution of Vessel Traffic Services to safe coexistence between automated and conventional vessels, *Maritime Policy & Management*. DOI: 10.1080/03088839.2021.1937739
- Ringbom, H., Røsæg, E. & Solvang, T. (2021) *Autonomous Ships and the Law*. 1<sup>st</sup> edn. Oxford: Routledge.
- Rosenfeld, L., Morville, P. & Arango, J. (2015) *Information Architecture*. 4<sup>th</sup> edn. Sebastopol: O'Reilly Media.
- Rosset, C. (2020) Turing-NLG: A 17-billion-parameter language model by Microsoft, *Microsoft Research Blog*, February 13. Available at: <https://www.microsoft.com/en-us/research/blog/turing-nlg-a-17-billion-parameter-language-model-by-microsoft/> (Accessed: 25 May 2022).
- Rutledal, D. Relling, T. & Resnes, T. (2020) It's not all about the COLREGs: a case-based risk study for autonomous coastal ferries, *3rd International Conference on Maritime Autonomous Surface Ship, MASS 2020*. Ulsan, November 11-12, 2020. IOP Publishing, p. 012016.
- Rødseth, Ø. J. & Burmeister, H. C. (2015a) *D10.1: Impact on Short Sea Shipping*. (MUNIN-report 21.09/2015). Available at: <http://www.unmanned-ship.org/mumin/wp-content/uploads/2015/10/MUNIN-D10-1-Impact-on-Short-Sea-Shipping-MRTK-final.pdf> (Accessed: 1 June 2022).
- Rødseth, Ø. J. & Burmeister, H.C. (2015b) Risk Assessment for an Unmanned Merchant Ship, *International Journal on Marine Navigation and Safety of Sea Transportation*, 9(3), pp. 357-364. DOI: 10.12716/1001.09.03.0
- Rødseth, Ø. J. (2018) Assessing business cases for autonomous and unmanned ships. In Technology and Science for the Ships of the Future, *Proceedings of NAV 2018: 19th International Conference on Ship & Maritime Research*. Trieste, June 20-22, 2018. Amsterdam: IOS Press, pp. 1033-1041.
- Smith, S. *et al.* (2022) *Using DeepSpeed and Megatron to Train Megatron-Turing NLG 530B, A Large-Scale Generative Language Model*. Available at: <https://arxiv.org/abs/2201.11990> (Accessed: 6 June 2022).
- Solaiman, I. *et al.* (2019) *Release Strategies and the Social Impacts of Language Models*. Available at: <https://arxiv.org/abs/1908.09203> (Accessed: 14 May 2022).
- Spradley, J. P. (2016) *Participant Observation*. Long Grove: Waveland Press.
- Stanton, N. A. *et al.* (2005) *Human Factors Methods: A Practical Guide for Engineering and Design*. 2<sup>nd</sup> edn. Burlington: CRC Press.
- Stickdorn, M. *et al.* (2018) *This Is Service Design Methods*. Sebastopol: O'Reilly Media.
- Turing, A. M. (1950) Computing Machinery and Intelligence, *Mind*, 59(236), pp. 433-460. DOI: 10.1093/mind/LIX.236.433
- Tvete, H. A. (2015) *The ReVolt: A new inspirational ship concept*. Available at: <https://www.dnv.com/technology-innovation/revolt/> (Accessed: 6 June 2022).



- Vaswani, A. *et al.* (2017) Attention Is All You Need, *Proceedings of Advances in Neural Information Processing Systems, NIPS 2017*. California, December 4-9, 2017. San Diego: Neural Information Processing Systems Foundation, pp. 5999-6009.
- Veitch, E. & Alsos, O. A. (2022) A systematic review of human-AI interaction in autonomous ship systems, *Safety Science*, 152, p. 105778. DOI: 10.1016/j.ssci.2022.105778
- Weizenbaum, J. (1966) ELIZA—a computer program for the study of natural language communication between man and machine, *Communications of the ACM*, 9(1), pp. 36-45. DOI: 10.1145/365153.365168
- Wróbel, K., Gil, M. & Montewka, J. (2020) Identifying research directions of a remotely-controlled merchant ship by revisiting her system-theoretic safety control structure, *Safety Science*, 129, P. 104797. DOI: 10.1016/j.ssci.2020.104797
- Xiao, A. *et al.* (2021) *Scaling ASR Improves Zero and Few Shot Learning*. Available at: <https://arxiv.org/abs/2111.05948> (Accessed: 3 June 2022).
- Yara (2022) *Crown Prince and youths christen world's first emission-free container ship*. Available at: <https://www.yara.com/corporate-releases/crown-prince-and-youths-christen-worlds-first-emission-free-container-ship/> (Accessed: 2 May 2022).
- Zhang, S. *et al.* (2022) *OPT: Open Pre-trained Transformer Language Models*. Available at: <https://arxiv.org/abs/2205.01068> (Accessed: 26 May 2022).
- Zhang, X. *et al.* (2021) Collision-avoidance navigation systems for Maritime Autonomous Surface Ships: A state of the art survey, *Ocean Engineering*, 235, p. 109380. DOI: 10.1016/j.oceaneng.2021.109380
- Ziarati, M. *et al.* (2011) Communication and Practical Training applied in Nautical Studies, *International Maritime English Conference, IMEC 2011*. Constanta, October 10-14, 2011. Constanta: Editura NAUTICA, pp. 41- 51.

# Appendix

## Appendix 1:

### INTERVIEW GUIDE FOR OBSERVATION ON SHIP BRIDGE AND VTS

#### Questions for Ship officers:

- Do you allow us to take anonymised notes and pictures?
- How long have you worked as a mariner?
- What communication tools do you use?
- Which one do you use the most?
- Which one do you prefer?
- Why do you prefer this communication tool?
- Can you tell us about a difficult situation and how you handled it using communication tools?
- What are some informal methods of communication that you use?
- How do you usually judge other ships' intentions?
- When you are confused about another ship's intention what is the most effective protocol?
- Do you contact first or do you get contacted the most? What do you prefer?
- Are there messages that should be delivered to you without you pinging?
- How would the transition to more autonomy on sea impact you the most?
- What is most exciting about MASS? What is most frightening?
- What do you think of maritime messaging and chatting with other ships and shore?
- How does autodocking work?

#### Questions for VTS operators:

- Do you allow us to take anonymised notes and pictures
- How long have you worked as a VTS operator?
- What communication tools do you use?
- Which one do you use the most?
- Which one do you prefer?
- Why do you prefer this communication tool?
- Can you tell us about a difficult situation and how you handled it using communication tools?
- What are some informal methods of communication that you use?
- How do you usually judge ships' intentions?
- When you are unsure about a ship's intention what is the most effective protocol to resolve this confusion?
- Do you contact first or do you get contacted the most? What do you prefer?
- Are there messages that should be delivered to you without you pinging them?
- How would the transition to more autonomy on sea impact you the most?
- What is most exciting about MASS? What is most frightening?
- What do you think of maritime messaging and chatting with other ships and shore?

## Appendix 2:

### SYSTEM USABILITY SCALE SCORE DATA

	M	SD( $\sigma$ )	1	2	3	4	5	6	7	8	9
I think that I would like to use this system frequently	2.22	0.62	2	2	1	2	2	3	3	3	2
I found the system unnecessarily complex	1.33	1.05	0	1	4	1	1	2	1	1	1
I thought the system was easy to use	2.77	0.91	2	4	1	3	4	2	3	3	3
I think that I would need the support of a technical person to be able to use this system	0.66	0.94	0	0	1	1	0	3	0	1	0
I found the various functions in this system were well integrated	2	0.66	2	3	1	3	2	2	2	2	1
I thought there was too much inconsistency in this system	1.66	1.05	2	3	3	3	1	1	1	0	1
I would imagine that most people would learn to use this system very quickly	3.22	0.78	3	4	2	2	4	4	3	3	4
I found the system very cumbersome to use	0.66	0.81	0	0	2	2	0	1	0	1	0
I felt very confident using the system	2	0.47	2	2	1	2	2	2	3	2	2
I needed to learn a lot of things before I could get going with this system	0.88	0.73	0	0	2	1	1	2	1	0	1
<b>SUS Score</b>	<b>67.5</b>		72.5	77.5	35	60	77.5	60	77.5	75	72.5