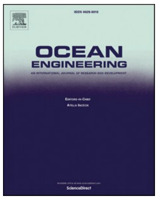
Ocean Engineering 310 (2024) 118641

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng



Research paper

Conversational user interfaces for maritime autonomous surface ships

Philip Hodne[a], Oskar K. Skåden[a], Ole Andreas Alsos[b,*], Andreas Madsen[c], Thomas Porathe[b]

[a] DNV AS, Høvik, Norway
[b] Department of Design, NTNU, Trondheim, Norway
[c] Department of Ocean Operations and Civil Engineering, NTNU, Ålesund, Norway

ARTICLE INFO

Index terms:
Autonomous ships
Human-AI collaboration
Maritime communication
VHF
Conversational AI
Human-machine interface

ABSTRACT

We propose integrating conversational user interfaces (CUIs) into Maritime Autonomous Surface Ships (MASS) to facilitate direct communication with other ships without human intervention. The aim is to enable other ships and shore stations to inquire about and discuss the autonomous vessel's status, decisions, and future intentions in natural language, offloading VHF radio conversations for MASS operators. Recent advancements in Natural Language Processing (NLP) technology show promise. This paper presents an experimental study conducted in a ship simulator using a functional prototype CUI system. While the CUI system was capable of providing coherent answers to questions, the results indicate that interacting with the system is less trustworthy than interacting with a human ship officer. Key challenges contributing to this included: (1) a missing connection between the CUI and an autonomy system, leading to inaccurate and incorrect answers, (2) speech that lacked intonation and human-like qualities, and (3) a perceived inability of the system to collaborate. The paper discusses several aspects of utilising CUI systems in a maritime shipping environment and concludes by proposing an approach to develop and implement them safely.

1. Introduction

In fog and darkness on the night of December 13th, 2021, the fatal collision between Scot Carrier and Karin Høj took place off the coast of Ystad, Sweden. The two seafarers aboard the Danish cargo ship Karin Høj were pronounced dead after the ship capsized in the wake of the collision. The much larger Scot Carrier on Karin Højs port side defied the International Regulations for Preventing Collisions at Sea (COLREGs) by suddenly turning starboard towards Karin Høj. The officer on watch on board Scot Carrier was later found guilty of involuntary manslaughter in Danish courts (Dixon, 2022). One of the crucial factors to the tragic event was that the ship officer, who admitted to drinking on the bridge, acknowledged that he did not respond to Karin Høj's requests over VHF radio (MAIB, 2023).

Through the development of autonomous ships, the maritime industry aims to enhance its safety, efficiency, and environmental impact, particularly in the short-sea shipping segment (IMO, 2021; Kim and Schröder-Hinrichs, 2021; Mørkrid et al., 2023; Rødseth, 2018). The successful autonomous navigation and docking of a fish feed carrier in Norwegian waters demonstrate how these benefits may soon be realized (Wennersberg, 2023). As newbuilds become more automated using advanced sensors and Artificial Intelligence (AI) technology, ship officers are expected to monitor multiple Maritime Autonomous Surface Ships (MASS) remotely from land-based Remote Operations Centres (ROCs) (Dybvik et al., 2020; MacKinnon et al., 2015), with a goal of being passive entities characterized by supervision and pre-emptive actions (Porathe et al., 2020). To achieve this, humans must trust the autonomous systems, which requires addressing the challenge of automation transparency (Porathe, 2019, 2021a). While previous research has focused on the awareness and confidence of operators in the ROC (Porathe et al., 2014; Ramos et al., 2019), this paper focuses on establishing trust with stakeholders outside of MASS operations, such as nearby crewed ships and shore stations. Addressing the human factors challenge of automation transparency is critical to achieving the goal of making MASS as safe as or safer than human ship officers. Lack of such transparency is expected to be one of the prominent risk factors in MASS operations (Chang et al., 2020; Veitch and Alsos, 2022).

2. Background

In this section, we present a pressing automation transparency challenge for MASS: the ability to communicate with nearby ships

---

* Corresponding author.
E-mail address: ole.alsos@ntnu.no (O.A. Alsos).

https://doi.org/10.1016/j.oceaneng.2024.118641
Received 24 June 2023; Received in revised form 9 June 2024; Accepted 1 July 2024

without the intervention of an onboard or land-based operator. We explain how it is critical for MASS to communicate independently through VHF radio and introduce our proposed solution.

2.1. Oceangoing robots without communication skills

Today's autonomous ships communicate their decisions and intent to nearby crewed ships predominantly through changes in course and speed or through an onboard or land-based operator. Without a human operator to handle radio communication, it is challenging for surrounding crewed ships and shore stations to negotiate and resolve ambiguous and dangerous situations with MASS. Information about the MASS's decisions and intentions is only available to the ROC operators (Alsos et al., 2022). Previous research has sought to resolve this. For example, the proposal for real-time route exchange on crewed ships' ECDIS systems would broadcast the MASS' intended route to surrounding vessels (Porathe and Rødseth, 2019; Porathe et al., 2020). Displays attached to the superstructure of vessels communicating status, decisions, and intent through text, colour and animation have also been proposed (Simic and Alsos, 2023; Porathe, 2021b).

While these contributions help increase trust in the autonomous system, the communication is one-sided. Ship officers and Vessel Traffic Service (VTS) operators also need to directly discuss and clarify intentions with the vessels to avoid collisions. Today, crewed ships primarily communicate through VHF radio, following expectations and regulations. The current strategy for MASS involves relaying that communication to the ROC, with the operator continuously monitoring the radio. However, as the operators may supervise multiple ships, monitoring multiple radios remotely may increase the operators' cognitive load and lead to a lack of situation awareness (SA), thereby increasing safety risks (Dybvik et al., 2020). Confusing answers, or worse, an absence of them can have detrimental consequences.

2.2. VHF radio communication

VHF radios are essential for ship-to-ship and ship-to-shore communication, required on 98.91% of the global merchant fleet (IMO, 2023, p. 18). Vessels and onshore entities like VTS, port authorities, and coast guards use it to collaborate, operating within a line-of-sight range of 50–70 nautical miles. Operating on preselected channels within multiple frequencies, they facilitate the exchange of crucial navigational information for collision avoidance and emergency responses.

2.3. Challenges with VHF communication

Although essential, VHF radio communication faces challenges such as static noise from environmental factors and interference, congested channels, and human factors like language barriers and operator inexperience and fatigue. This is exacerbated by messages being broadcasted to everyone in the radio's range, even if they are intended for one ship only. These issues pose a significant risk of misunderstandings and collisions (Valle, 2011; IALA, 2022; Porathe et al., 2014; Ziarati et al., 2011). No advanced systems for transcribing or translating VHF messages are widely implemented yet (Nakilcioğlu et al., 2023; Gözalan et al., 2020; John and Reimann, 2021), making constant monitoring and regulations on how to speak a need. Most prominent is the Standard Marine Communication Phrases.

2.4. Standard Marine Communication Phrases (SMCP)

Introduced by the IMO, the SMCP standardizes English phrases for maritime communication, covering navigation, safety, and operational communication. The general rules for initiating contact are illustrated in Fig. 1 as an example. Its use is mandatory for international voyages (IMO, 2001), with the phrases aiming to improve clarity and efficiency. However, usage often deviates due to time pressures and stress in real-world situations (Böstrom, 2020; Dževerdanović-Pejović, 2013). Informal language and the use of native languages are common.

3. Conversational user interfaces for MASS

With the aim of offloading VHF radio communication for ROC operators, we propose a novel system that utilizes a Conversational User Interface to share intentions and resolve ambiguous or dangerous situations without human intervention.

3.1. Conversational User Interfaces

Conversational User Interfaces (CUIs) are human-machine interfaces that simulate human-like conversations, enabling users to interact with computer systems and services using spoken or written language (McTear, 2002). CUIs are designed to emulate human dialogue and aim to be an intuitive method of collaborating with machines. They are often built as systems following a six-step process, as illustrated in Fig. 2.

CUIs rely on Natural Language Processing (NLP), a field that employs statistical learning to understand and process human language. Deep learning techniques, including notably the Transformer architecture (Vaswani et al., 2017), have demonstrated exceptional performance in various NLP tasks such as speech recognition, language understanding, and synthesis (LeCun et al., 2015). These advances have been fuelled by the availability of large datasets and increased computational power. This paper leverages these modern NLP techniques aiming to demonstrate the viability of a CUI for maritime communication.

3.2. Applying CUIs to the domain of MASS

While CUIs are well-established in other domains such as customer support (McTear et al., 2016, pp. 11–16), their application in the maritime shipping sector is new. Given the safety critical nature of shipping, developers would have to ensure that CUIs are robust and context-aware, to prevent serious consequences. In this context, it is essential to address the constraints posed by MASS operations and VHF radio communication. MASS relies on sensor data, including AIS, radar, lidar, and computer vision, to create a comprehensive model of the surrounding environment through sensor fusion. This *model of the world* aids the vessel in understanding its position relative to other objects such as land or nearby vessels, and the autonomy system determines the appropriate navigation actions (illustrated in Fig. 3). As messages through VHF have no metadata, it can be challenging to identify the sender of the messages directly. To effectively communicate, the CUI must utilize the world model, accurately identify the user, and provide contextually relevant explanations aligned with their understanding. It

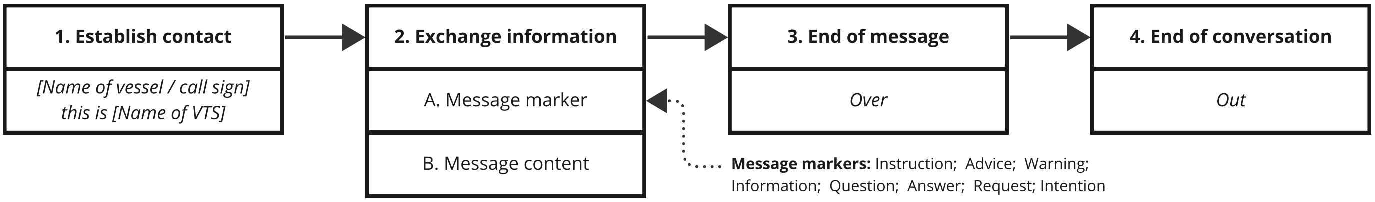


Fig. 1. Guidelines for SMCP sentence structuring (IALA, 2022).

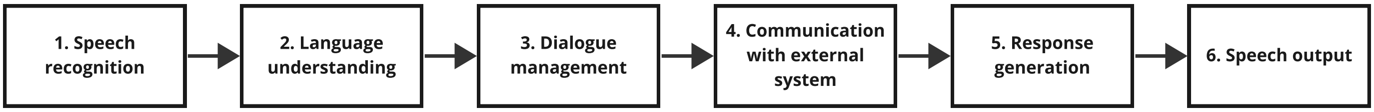


Fig. 2. The six-step process of conversational user interfaces (McTear, 2002).

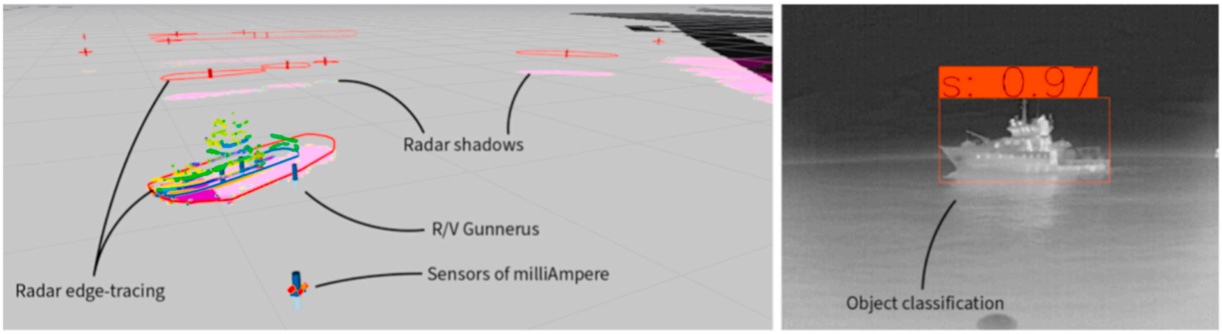


Fig. 3. From the sensor fusion of NTNU's autonomous ferry MilliAmpere (with permission from Brekke et al., 2022).

should recognize when it's being contacted and adapt the conversation to the mental model of the ship officer or operator.

Next, the CUI system must hold coherent conversations based on available data from its autonomy system, explaining decisions and potential outcomes. Understanding the semantics of the conversation, the system's aim is to offer reasoned answers regardless of phrasing. Crucially, the system should be able to collaborate with stakeholders through dialogue to jointly decide the next steps. This should be done while adhering to established rules of communication, notably the SMCP. The system should also work in tandem with the navigation system to execute agreed-upon steps. Based on the conversation, navigational commands would be interpreted and executed in ways that are physically possible and appropriate in the real world. Trust is established by balancing efficiency, safety, and adherence to sea rules such as COLREGs when translating these into movement.

Finally, the system includes a ledger of exchanged messages for ROC operators and the ability to override intended actions, ensuring control. Incorporating redundancy as a safety measure, the system alerts ROC operators when necessary, and enables direct contact with the operator during emergencies for stakeholders interacting with the CUI. See Fig. 4 for an overview.

Evidence that conversational user interfaces can assist in maintaining situation awareness in the maritime domain have been established, amongst others, between operators and Autonomous Underwater Vehicles (Robb et al., 2018, 2019). As demonstrated in their paper on grounding language in robotic affordances, Ahn et al. (2022) presents a robot that executes several tasks in its environment based on complex and temporally extended instructions, while Vemprala et al. (2023) highlighted the successful translation of natural language action sequences into executable code for drone operations in a simulator.

3.3. A system not limited to VHF

We have proposed that the CUI should use VHF radio technology to

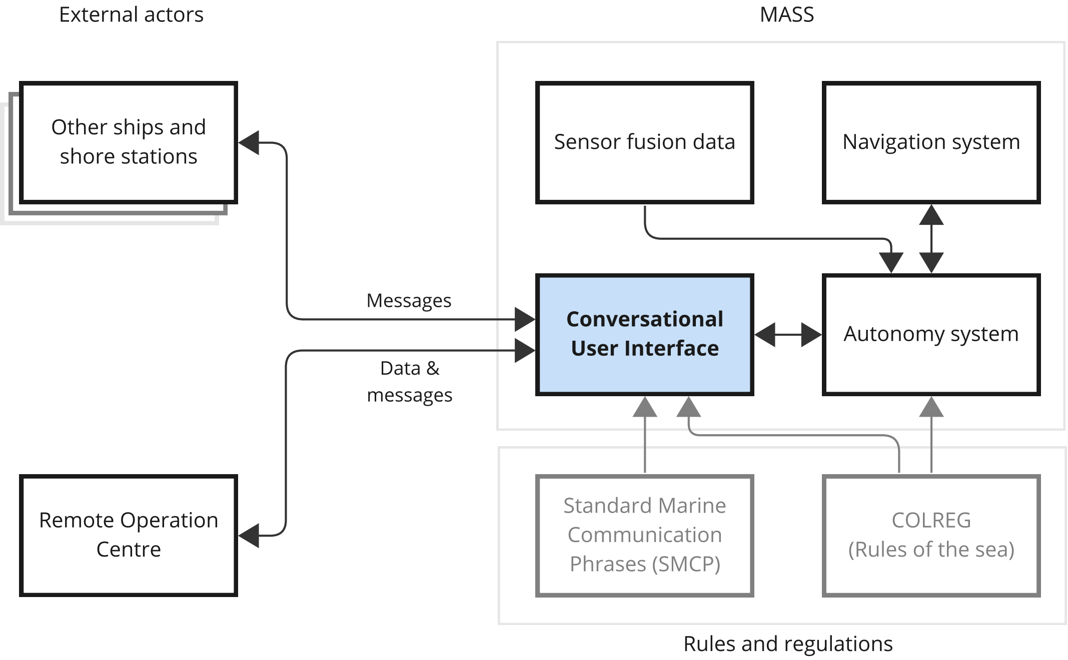


Fig. 4. The Conversational User Interface system in relation to systems onboard the MASS, external actors, and rules and regulations.

send and receive messages. At least in the beginning. The CUI system is not inherently limited to a particular communication technology. However, pivotal to the success of MASS is their ability to integrate with established standards. As VHF radios are already adopted by virtually all merchant ships, its independent installation on the MASS facilitates easy rollout.

Still, under the IMO regulatory exercise for MASS, it was stated that developing new instruments will likely be the most appropriate way to go for radio communications (IMO, 2021, p. 30). Considering the challenges associated with VHF, CUI systems should be designed to work independently of communication technology. Future instruments may even offer additional functionality, enhancing automation transparency between humans and autonomous vessels. Thus, if another communication technology is introduced, the CUI system should seamlessly adapt and integrate simultaneously with these. Looking ahead, in a traffic setting where only MASS' are present, one can imagine a communication system devoid of human voices, relying solely on machine-to-machine interaction to navigate encounters, as outlined by Akdağ et al. (2022). However, in a mixed traffic environment where conventional crewed ships coexist, enabling automation transparency is still crucial. Facilitating the use of CUI systems would allow crewed ships in the vicinity to listen in and develop a shared understanding of the ongoing situations, fostering a collective awareness.

4. A prototype CUI system

In this section, we present a proof-of-concept CUI system intended for VHF that we evaluated in a ship simulator. We outline the technologies used in the study to build the functional CUI system. We first present the NLP models, before describing how we combined these to develop the system itself.

4.1. NLP models

4.1.1. Whisper

Trained on 680 000 h of speech, Whisper is a transformer-based automatic speech recognition model that approaches human-level robustness and accuracy in speech recognition in several languages (Redford et al., 2022). It is a multitasking model that can perform multilingual speech recognition, speech translation, and language identification. This means the system is trained to understand several languages, various accents, speech obscured by background noise, and technical idioms, and translate these into most languages. The model is open source and can be accessed in a standalone version and through a hosted API service. It includes several model sizes (tiny, base, small, medium, and large) each offering speed and accuracy trade-offs.

4.1.2. GPT-3.5

GPT-3.5, an upgraded version of the third-generation Generative Pre-trained Transformer (Brown et al., 2020), is a generative language model trained on vast amounts of internet data. It excels in generating coherent and high-quality text output for various tasks such as question-answering, summarization, translation, and creative writing. It is proficient in semantic comprehension, irrespective of grammatical errors. The model has a context window for holding conversations spanning multiple messages. While it produces high-quality results, it faces challenges such as generating fictitious information (known as LLM hallucination) and exhibiting biases and discriminatory language. To mitigate these issues, reinforcement learning from human feedback (RLHF) has been employed (Ouyang et al., 2022). The technique allows the use of human feedback to optimize a model's policy. GPT-3.5 is not publicly available and can only be accessed through its API.

4.1.3. Neural2

Neural2 is a text-to-speech model which converts text into human speech. It was made available as an API in 2022 (Jia and Cattiau, 2021). The underlying technology, PnG NAT, is the result of a merger of two prior models (Jia et al., 2021; Shen et al., 2021), giving improved prosody and pronunciation in the synthesized speech. Text can be synthesized as a specific voice across multiple languages through distinct API endpoints.

4.2. CUI system architecture

The CUI system architecture is presented in Fig. 5. The CUI system prototype was developed using a Python backend and a React.js frontend. It received recorded sound files as input and generated spoken answers. Pressing the Enter key activated the microphone in the frontend, allowing users to record their messages. These recordings were then sent to the backend for transcription using Whisper, utilizing the multilingual Base model for efficiency. The transcribed text was forwarded to GPT-3.5 through the OpenAI API, with each request using an identical prompt that asked for information from a list of past decisions, current status, and future intent (Appendix A). Inspired by the Retrieval-Augmented Generation method for NLP tasks (Lewis et al., 2020), this prompt served as the CUI system's model of the world and significantly influenced the answers provided. To minimize randomness, the model's outputs were configured to be more deterministic. Each API call was independent, disregarding previous questions and answers to avoid exceeding the model's context window limitations. The text response from GPT-3.5 was then sent to the Google Cloud Text-to-Speech API using the Neural2 British male voice endpoint (en-GB-Neural2-B). The API returned an OGG format sound file encoded in BASE64, which was decoded in the frontend and played audibly in the browser. Logs were kept for Whisper, GPT-3.5, and Neural2 inputs and outputs. Although the system was limited to English, the official language at sea, it aimed to understand various accents and dialects to accommodate the diverse range of accents and English proficiency commonly encountered.

5. Method

An experiment was designed to test the CUI system's ability to communicate with humans in a way that helps clarify a confusing situation. An overarching goal was set: the system should provide the same or an excess amount of trust compared to a human ship officer. In this section, we will present how we designed the experiment to test this.

5.1. Experimental design

The question was whether communicating through speech with a CUI system is perceived at least as trustworthy as interacting with a human ship officer (H0) or if there is a significant difference in perception (H1). The benchmark for this hypothesis was the Checklist for Trust Between People and Automation (Jian et al., 2000), which measures trustworthiness based on ability, integrity, and benevolence (Table 1). The checklist uses a 7-point Likert scale, where lower scores in the first five questions indicate higher trust, and higher scores indicate higher trust for rest.

The Checklist for Trust is a widely used self-report measure for trust in automated systems (Kohn et al., 2021), although potential positive bias has been reported (Gutzwiller et al., 2019). It could be argued that adding objective measures of trust could be beneficial. Subjective self-report measures have limitations in capturing the underlying construct of trust, as they rely on conscious awareness and accurate self-reporting. Objective measures like response times or neural activity could provide a more direct assessment of trust-related cognitive processes. However, subjective surveys offer rich qualitative data, flexibility, and depth of insight into trust, making them a feasible and cost-effective option within the scope of this study.

We tested the CUI system in a ship simulator, comparing its performance to that of a human ship officer in an identical scenario. Making the experiment blind aimed to prevent participants from preparing

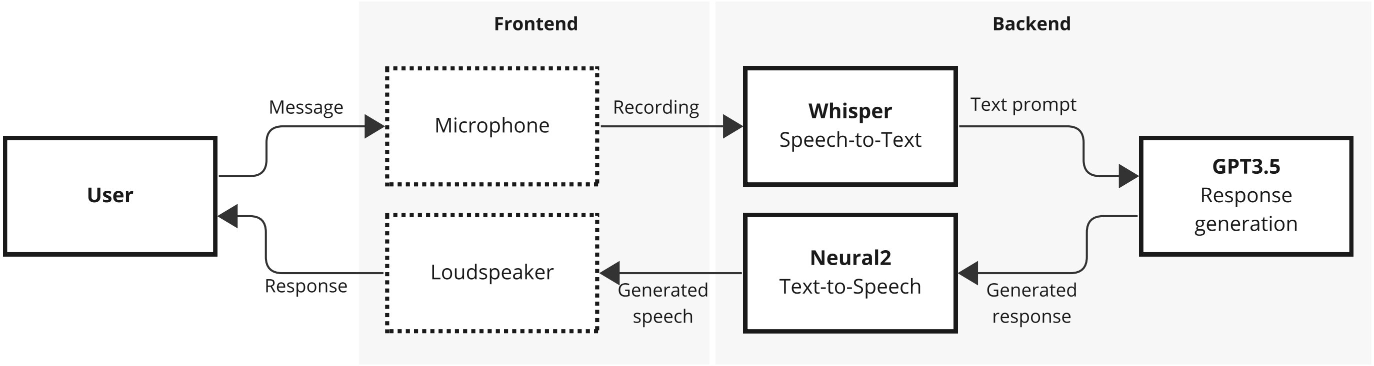


Fig. 5. System architecture for the CUI system.

Table 1
Checklist for trust between people and automation.

Order	Question
1	The system is deceptive.
2	The system behaves in an underhanded manner.
3	I am suspicious of the system’s intent, action, or outputs.
4	I am wary of the system.
5	The system’s actions will have a harmful or injurious outcome.
6	I am confident in the system.
7	The system provides security.
8	The system has integrity.
9	The system is dependable.
10	The system is reliable.
11	I can trust the system.
12	I am familiar with the system.

themselves in advance and thus allowed us to observe their natural reaction. We controlled for variables such as the simulator apparatus, scenario, brief, and procedure, with the only independent variable being the speech output in response to VHF questions. The intervention group would get a response from the CUI system, while the control group would speak to a human ship officer. Testing in a ship simulator, the goal of the apparatus, procedure, and simulator case was designed to promote natural behaviour by participants while minimizing unforeseen actions.

5.2. Participants

A total of 36 participants, who were first- or second-year nautical students at a vocational college, were recruited for the study. All participants were native Norwegians. The participants had to meet two criteria: (1) being enrolled as nautical students at the college and (2) having appropriate background training and experience, including at least 6 months of sea service in lower rank positions and a minimum of 60 h of simulator training.

The participants were randomly selected from both classes as a single population and then assigned randomly into either the intervention or control group. The intervention group consisted of 19 participants whose VHF calls were answered by the CUI system. The control group comprised 17 participants whose calls were handled by a human ship officer from the college staff located in a nearby ship bridge simulator. The participants in the intervention group had an average seagoing experience of about 3 years, while the control group participants had an average of about 2 years (See Table 2). Experience might influence comfort with procedures and trust in communication protocols, but anonymization prevented us from exploring this aspect in our data to see if age or experience had any bearing on the response.

Table 2
Months of experience for intervention and control group.

Group	Mean	Median	SD	Min	Max
Intervention	38.3	36	31.3	6	140
Control	21.6	12	26.6	12	120

None of the participants were excluded based on the screening or altering their ship’s course. With five questionnaires incorrectly filled out and discarded, the total sample size for the Checklist for Trust between People and Automation ended up at 31. The same officer was assigned to all control group participants and was given instructions on how to respond. The tests were conducted individually, one participant at a time.

5.3. Apparatus

The experiment was conducted in a Kongsberg K-SIM Navigation Ship Bridge Simulator (Fig. 6) with a predefined scenario designed to discourage participants from overriding their own vessel’s route and encourage passive watchkeeping and clarification over VHF. The aim was to ensure more consistency in measurements and avoid unintended variables in the experiment.

Intervention group: The CUI system software was installed on a desktop computer hidden in a cabinet underneath the desk. The sound was recorded and played through a wired conference speaker, also hidden from view. The existing VHF radio on the bridge was disconnected, and a custom-made button was embedded into another identical VHF radio to enable the button press on the radio to be picked up as an Enter keypress on the computer. Thus, pressing the VHF call button caused speech from the room to be recorded by the hidden conference speaker. The objective was to create an experimental environment that closely approximated the operational environment of ship bridges, while minimizing exposure to hazardous real-life situations and affording greater control over dependent and independent variables. Note that the CUI system was not connected to any data from the simulator and would answer requests based on its prompt. Also, any messages recorded would be answered even if not directed at the CUI system’s ship.

Control group: The research apparatus and protocol for the control group were identical except for two points: (1) instead of responses from the CUI system, the ship officer would answer the participants from a sister simulator using the available simulator equipment, and (2) they would not participate in the debrief interview. The ship officer was informed of the ship’s sailing route, and instructed not to intervene with the ship’s sailing, and only respond to VHF calls from the students. It was not possible to mitigate a “learning bias”, meaning that the ship officer could improve performance after each participant. However, this was addressed through a predefined script that the ship officer was to follow (Appendix B).

Fig. 6. The Ownship's bridge from inside the simulator used. The surrounding waters are showing on the screens.

5.4. Procedure

The experimental procedure consisted of the following 4 parts where the participants (1) filled out an informed consent form and received an experimental brief, (2) performed a navigation scenario in the simulator, (3) filled out a post-experiment questionnaire, and (4) participated in a debrief interview.

The experiment began with the participants being welcomed into the main room, where they handed in their signed consent form. They were then led into a separate room and given the case brief, read out loud from a document (Appendix B). No follow-up questions were answered. Next, they were taken to the simulator room, where the navigation scenario was started by the facilitators. Video cameras with sound were used to observe them in real-time. There were no interactions with the facilitators during the simulator sessions. Results were discarded if participants altered the outcome of their ship. After the simulation ended, they filled out the Checklist for Trust between People and Automation. Questionnaires were discarded if filled out incorrectly. After filling out the form, a debrief qualitative interview was conducted in the main area using a structured guide with 5 open-ended questions (see Appendix C).

5.5. Simulator case scenario

The case scenario was based on the collision between *Scot Carrier* and *Karin Høj* as described in the introduction (MAIB, 2023). Participants controlled the Ownship, a suezmax oil tanker named *Antares*, while a MASS named *Atlantic* acted as the Targetship. Atlantic followed a route similar to Scot Carrier but avoided a collision by turning to starboard, slowing down, and waiting for Antares to pass before navigating behind it (See Fig. 7 for reference). At the beginning of the scenario, the participants observe on both radar and AIS that a ship is overtaking them from their port stern. After 4 min, they have been overtaken, and the Targetship changes its course to starboard to pass ahead of the Ownship.

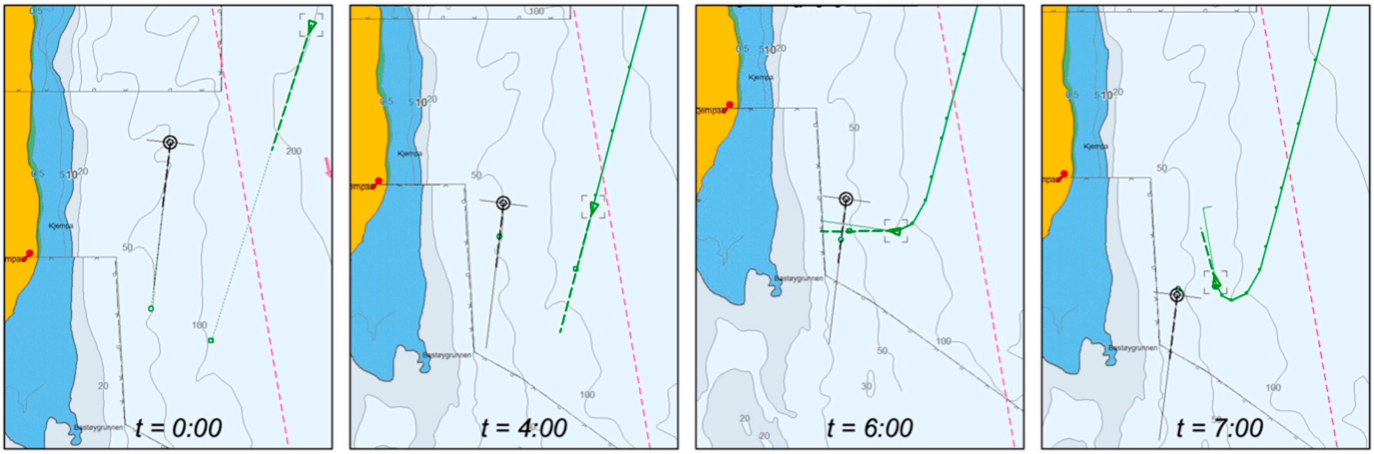


Fig. 7. The scenario depicts the Target Ship overtaking Ownship and subsequently turning starboard aft of it. The time is measured in minutes from the start of the scenario.

The CPA is 0 nm, and TCPA is 3 min. Due to the Ownship's stopping distance, and the depth limitations on the ship's starboard side, the participants have no other means than communication with the target ship to avoid a collision.

5.6. Data collection and analysis

In addition to observing the participants, all sessions and subsequent debrief interviews were recorded and transcribed. The questionnaires were coded by hand. To create a singular trust score for the Checklist of Trust, the first five questions were score-reversed (using *8-X*). The total sum for each participant, as well as mean averages and standard deviations for each question, were calculated. A two-tailed, independent samples *t*-test was used for each question to determine whether the mean difference between the intervention and control groups was statistically significant. Alpha levels (α) were set to 0.05, 0.01 and 0.001. Assumption tests for normality and variance homogeneity were conducted using Shapiro-Wilks and Leneve's tests respectively.

6. Results

This section presents the findings from the simulator test, the results from the Checklist of Trust questionnaire, and feedback from the debrief interviews.

6.1. Intervention group observations

Based on our observations of the simulator scenarios, we noticed that all the participants responded appropriately. Once they realized that the overtaking by the Targetship posed a risk, they promptly contacted the ship to gather information about its intentions. All students inquired about the intent differently. Some diligently followed SMCP, while others were more informal. A few seconds later the CUI system answered them. For the intervention group participants, the experiences were dichotomous. When the system answered accurately, there was little confusion and few follow-up questions:

> ***Participant:*** *Atlantic, Atlantic. This is Antares. I am wondering what is your intentions?*
>
> ***MASS:*** *This is Atlantic. We are currently on a south heading, at 12 knots. We are aware that we are on a collision course with you and intend to slow down and navigate astern of you.*
>
> ***Participant:*** *Ok. Thank you, thank you.*

Here, the participants showed few signs of worry and continued with watchkeeping until the simulation ended. Although the answers provided by the CUI were coherent and direct for some participants, for most of them, the responses were ambiguous and unclear, resulting in extensive back-and-forth communication with repetitive answers from the CUI. It often failed to understand or properly respond to certain requests, tending to generate fictitious information in parts of messages (known as LLM hallucination). For example, many responses contained incorrect information, such as confusing starboard and port, and referring to itself as a ship officer, which did not accurately represent the situation.

> ***MASS:*** *This is Atlantic. We are currently headed south, 197°, at a speed of 12 knots. We are on your starboard side and are turning starboard towards you. We are on collision course and intend to slow down and navigate astern of you.*
>
> ***Participant*** (to himself): *Starboard side?*

Responses generated based on a static prompt sometimes contained outdated information, depending on when participants established contact. Additionally, some participants experienced distress due to the CUI system's inability to collaborate and lack of firm intent.

> ***Human:*** *Atlantic. Antares. Please change course so we do not have a too small CPA.*
>
> ***MASS:*** *I am not able to change course. However, I intend to slow down and navigate astern of you to avoid a collision.*
>
> ***Participant:*** *You will reduce speed, thank you.*
>
> ***MASS:*** *I apologize for not being able to meet you. I am currently on a course heading south at 12 knots, and I'm unable to change course at this time. I intend to slow down and navigate astern of you to avoid a collision.*

Upon evaluating the conversations, we also noticed a recurring pattern: poor inputs led to poor outputs. In cases where participants failed to deliver VHF messages clearly or intelligibly, it had a negative impact on the quality of the conversation. The system tended to provide excessively detailed and lengthy responses, failing to comprehend the flawed input, and neglecting to seek clarifications. It responded to transcribed inputs indiscriminately without considering potential errors or ambiguities (See Fig. 8).

6.2. Checklist for Trust between People and Automation

There was a clear disparity in the perceived trust reported by participants in the intervention and control groups. Fig. 9 illustrates that the CUI system received lower scores overall compared to the ship officer. A dichotomy within the intervention group is also present. While some participants consistently rated the CUI highly across all 12 questions, others consistently rated it poorly. In contrast, the control group showed less variation, with most scores falling between 6 and 7 for all questions. This inconsistency can be seen in the total sum of scores for each participant.

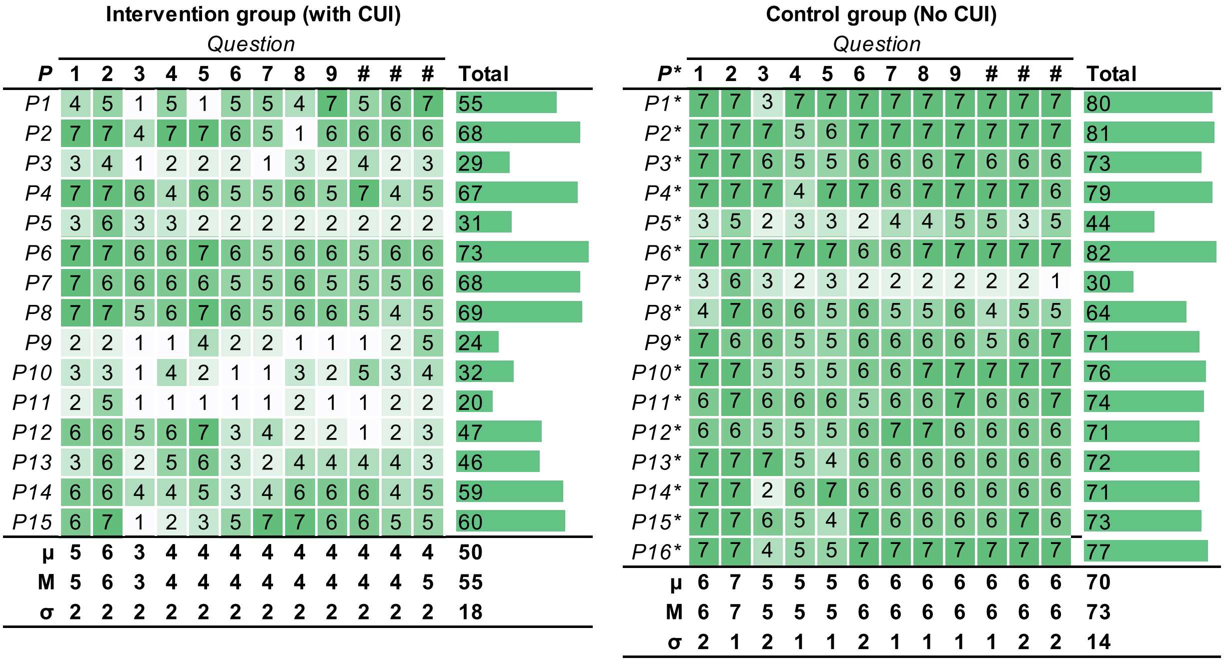
As for *t*-test assumptions, the Shapiro-Wilks and Levene's tests reveal normal data distribution for most questions, but unequal variance between groups respectively (Tables 3 and 4). The *t*-test indicated significantly lower trust levels in the CUI system across all dimensions compared to human ship officers, with significant differences observed in all except 4 and 5 (Fig. 10). These findings support rejecting the null hypothesis, suggesting participants perceived the CUI system as less trustworthy compared to interacting with a human ship officer.

6.3. Debrief interviews

During the debrief interview, most discussions focused on the effectiveness of communication, the potential for improvement in clarity and transparency, as well as the use of SMCP communication descriptors (such as "Question" and "Answer") to better facilitate understanding and action. Participants emphasized that the main reason for their perceived lack of trust in the CUI system was its inconsistency between words and actions. They were concerned about the system's understanding of directions, position, and ship names. The participants stressed the importance of the CUI system delivering on its promises. Consequently, they expressed uncertainty about the system's ability to handle complex conversations. Although they recognized the computer-generated voice, they found it easy to comprehend. However, the lack of intonation and emotional expression in the speech contributed to their mistrust. Participants mentioned that they could easily discern that the CUI was not a real human, which led to some scepticism.

Fundamentally, the students had few problems with a computer answering them over VHF – given that the answers were accurate and reliable. While most of the students were able to easily comprehend the CUI system, a majority still expressed scepticism towards the use of the technology in operation. This scepticism was present even in many of the participants who had a satisfactory conversation. The reasoning varied, but several worried about reliability in more complicated situations. Talking to a CUI in large, open water was considered less problematic than implementing it in congested and more heavily trafficked waters. Still, some reflected that trust in a CUI is likely to improve once

Fig. 8. From a simulator session. The CUI system responded promptly, but the answers were often confusing.



Intervention group (with CUI)

Question

P	1	2	3	4	5	6	7	8	9	#	#	#	Total
P1	4	5	1	5	1	5	5	4	7	5	6	7	55
P2	7	7	4	7	7	6	5	1	6	6	6	6	68
P3	3	4	1	2	2	2	1	3	2	4	2	3	29
P4	7	7	6	4	6	5	5	6	5	7	4	5	67
P5	3	6	3	3	2	2	2	2	2	2	2	2	31
P6	7	7	6	6	7	6	5	6	6	5	6	6	73
P7	7	6	6	6	6	5	5	6	5	5	5	6	68
P8	7	7	5	6	7	6	5	6	6	5	4	5	69
P9	2	2	1	1	4	2	2	1	1	1	2	5	24
P10	3	3	1	4	2	1	1	3	2	5	3	4	32
P11	2	5	1	1	1	1	1	2	1	1	2	2	20
P12	6	6	5	6	7	3	4	2	2	1	2	3	47
P13	3	6	2	5	6	3	2	4	4	4	4	3	46
P14	6	6	4	4	5	3	4	6	6	6	4	5	59
P15	6	7	1	2	3	5	7	7	6	6	5	5	60
μ	**5**	**6**	**3**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**50**
M	**5**	**6**	**3**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**5**	**55**
σ	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**18**

Control group (No CUI)

Question

*P**	1	2	3	4	5	6	7	8	9	#	#	#	Total
*P1**	7	7	3	7	7	7	7	7	7	7	7	7	80
*P2**	7	7	7	5	6	7	7	7	7	7	7	7	81
*P3**	7	7	6	5	5	6	6	6	7	6	6	6	73
*P4**	7	7	7	4	7	7	6	7	7	7	7	6	79
*P5**	3	5	2	3	3	2	4	4	5	5	3	5	44
*P6**	7	7	7	7	7	6	6	7	7	7	7	7	82
*P7**	3	6	3	2	3	2	2	2	2	2	2	1	30
*P8**	4	7	6	6	5	6	5	5	6	4	5	5	64
*P9**	7	6	6	5	5	6	6	6	6	5	6	7	71
*P10**	7	7	5	5	5	6	6	7	7	7	7	7	76
*P11**	6	7	6	6	6	5	6	6	7	6	6	7	74
*P12**	6	6	5	5	5	6	7	7	6	6	6	6	71
*P13**	7	7	7	5	4	6	6	6	6	6	6	6	72
*P14**	7	7	2	6	7	6	6	6	6	6	6	6	71
*P15**	7	7	6	5	4	7	6	6	6	6	7	6	73
*P16**	7	7	4	5	5	7	7	7	7	7	7	7	77
μ	**6**	**7**	**5**	**5**	**5**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**70**
M	**6**	**7**	**5**	**5**	**5**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**73**
σ	**2**	**1**	**2**	**1**	**1**	**2**	**1**	**1**	**1**	**1**	**2**	**2**	**14**

Fig. 9. Trust scores for all participants, intervention group and control group.

Table 3
Shapiro-Wilks normality tests for the intervention and control group.

Intervention				Control group			
Question	Test Statistic	p-value	Normality ($p < 0.05$)	Question	Test Statistic	p-value	Normality ($p < 0.05$)
1	0.8176	0.0062	Yes	1	0.6059	1.8405e-05	Yes
2	0.8336	0.0102	Yes	2	0.5866	1.2129e-05	Yes
3	0.8210	0.0069	Yes	3	0.8595	0.0188	Yes
4	0.9197	0.1907	No	4	0.8811	0.0403	Yes
5	0.8566	0.0215	Yes	5	0.8971	0.0723	No
6	0.8813	0.0496	No	6	0.6802	0.0001	Yes
7	0.8667	0.0302	Yes	7	0.7263	0.0003	Yes
8	0.8846	0.0556	No	8	0.7290	0.0004	Yes
9	0.8502	0.0175	Yes	9	0.6406	4.0193e-05	Yes
10	0.8599	0.0241	Yes	10	0.7800	0.0015	Yes
11	0.8596	0.0238	Yes	11	0.7134	0.0002	Yes
12	0.9191	0.1867	No	12	0.6481	4.7757e-05	Yes

Table 4
Leneve’s variance homogeneity tests for the intervention and control group.

Test statistic	p-value	Decision
33.10	2.25e-08	Reject H0: Variances are not equal

people are exposed more to it; “It is natural to be sceptical of new technology”, said one of the participants.

7. Discussion

In this paper we sought to demonstrate the feasibility of

	Question													
Trust level	1	2	3	4	5	6	7	8	9	10	11	12	μ	σ
Intervention (CUI)	4,9	5,6	3,1	4,1	4,4	3,7	3,6	3,9	4,1	4,2	3,8	4,5	4,2	0,64
Control (human)	6,2	6,7	5,1	5,1	5,3	5,8	5,8	6,0	6,2	5,9	5,9	6,0	5,8	0,48
Difference	1,3	1,1	2	1	0,9	2,1	2,2	2,1	2,1	1,7	2,1	1,5		

Difference
α=0,05
α=0,01
α=0,001
Not significant

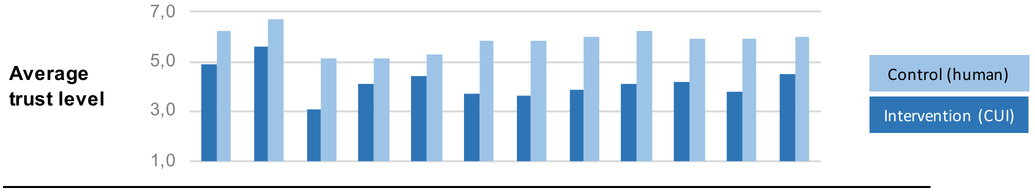


Fig. 10. Trust scores: average for all questions, significant differences between intervention group and control group.

incorporating conversational user interfaces (CUI) into Maritime Autonomous Surface Ships (MASS). Such a system would allow other ships and shore stations to communicate with them in natural language. In an experimental study we tested whether navigators would trust answers coming from our prototype CUI attempting to mimic a MASS.

7.1. On the trust of the CUI system

The results showed that there was a significant disparity in trust when talking to the CUI system, compared to that of a human ship officer. The self-report questionnaire reported lower total scores per participant and lower average question scores across all measurements of trust compared to the ship officer. Keeping in mind the violation of the assumption of homogeneity of variances for the *t*-test, results showing statistically significant differences, we argue that the null hypothesis can be rejected: talking to the prototype CUI system is not as trustworthy as talking to a human ship officer. At this juncture, the authors hold the viewpoint that employing the models and our system architecture for a CUI system in MASS would pose significant risks and, therefore, advocate for its avoidance. We believe there are three main drivers for this.

Frail world model: The lack of a robust, dynamic world model was the main issue affecting the CUI system. With a static, text-based prompt as the world model, the CUI struggled to provide contextualized answers, resulting in logical errors. Ship officers rely on multiple senses to gather information from the environment, such as looking out the bridge, checking ECDIS or RADAR displays, listening to engine sounds or alarms, and feeling the vibrations from the engine and waves against the hull. This highlights the necessity of integrating CUI systems deeply into autonomy and navigation systems. In safety-critical settings like maritime shipping, ensuring absolute accuracy in responses is crucial for the successful implementation of CUI systems, allowing no margin for error.

Monotone voice: The second problem was that the synthetic voice lacked natural intonation. Human communication is not solely based on the information exchanged but also on the nuances conveyed through vocal intonation, pitch, and rhythm. These elements provide cues about the speaker's emotional state, intentions, and level of confidence. In contrast, the CUI system's monotone and mechanical voice lacked these natural prosodic features, which may have led to reduced trust. The ability to rely on emotional cues to gauge sincerity and authenticity in communication is a crucial part of maritime communication.

Minimal collaborative capacity: Although not solicited through the simulator case, the ability to discuss and collaborate on how to navigate was not possible. We instructed through the prompt that the CUI should not allow any changes to its route. Still, maritime communication often involves discussions, negotiations, and confirmations, as well as collaboration on navigation strategies, adjustments to routes, and sharing of insights. Addressing this limitation is essential for the successful integration of CUI systems into maritime operations.

7.2. Limitations

There are several limitations to our study that may impact the results. Firstly, due to data anonymization, we could not measure the correlation between trust and individual factors, potentially missing important insights. For example, we couldn't determine if seagoing experience affects behaviour in the simulator. This could be mitigated in future similar experiments. Additionally, although the experiment was intended to be blind, participants recognizing the AI voice could have introduced a bias, influencing their behaviour or trust levels. To build on this, the CUI system was not tested in-situ. Rather, it was conducted in a ship simulator, and the "VHF" used was a high-quality conference speaker directly connected to the computer, providing high quality sound input and output. In contrast, the signal quality of real VHF messages is much lower. These factors impact the study's validity. As mentioned earlier, it is important to note that the accuracy of the transcription by the speech-to-text model may not be as reliable when dealing with true VHF radio signals. This limitation can potentially introduce inaccuracies in the collected data. On that note, no measure of word error rate on the input was used to assess the effectiveness. This way, we can't accurately determine if the lack of trust was influenced by answers generated on inaccurate inputs.

Furthermore, the case tested was narrow, testing mostly the system's ability for question answering. We concur with the participants that reflected on the ability of the current CUI system to answer coherently in more complex situations. Also, with a small sample size, the generalizability of the findings becomes limited. This was also suggested by the violations of the Levene's test and some of the assumptions of the Shapiro-Wilk tests. The participants involved in the study may not represent the diverse range of individuals who would interact with the CUI system in real-world maritime scenarios. This lack of diversity reduces the external validity of the study's conclusions. These limitations may have led to an unjustified higher trust in the CUI system and highlight the need for future research to address these concerns and conduct studies in more realistic environments with larger and more diverse participant samples.

Despite its limitations, this study arguably showcases the prospective viability of key CUI system functionalities for MASS. However, significant gaps remain that need to be addressed before implementing them in safety-critical settings. Three essential areas critical to the success of CUI systems have been identified: (1) Connecting world models to the CUI, (2) meeting dialogue expectations, and (3) translating conversations into movement. We will explore how to design these next.

7.3. Connecting world models to the CUI

The study's CUI prototype has been hindered by its reliance on text-based prompts, limiting its ability to comprehend and discuss the situation effectively. This limitation highlights the fact that it may not be the language model's fault for providing confusing answers. While language models possess impressive linguistic capabilities, without a comprehensive understanding of real-world dynamics that a robust world model provides, the answers are often lacking in depth and accuracy. Our prototype CUI thus restricted the model's capacity to comprehend and engage with the complexity of the situation. To address this, connecting the CUI to the MASS's autonomy system is crucial, enabling it to utilize dynamic data as a foundation. Recent multimodal model research (Girdhar et al., 2023) has shown promise in image generation by recognizing relationships across various modalities such as images, video, audio, text, depth, thermal, and inertial measurement units (IMUs). When a CUI can leverage the same senses as humans, it stands to gain a deeper understanding of the world and offer more informed and relevant responses. Additionally, a focus on explainable AI (Veitch and Alsos, 2022), where the system explains the data inputs used to generate its output and ensuring that the language models do not generate fictitious information, are vital for establishing trust and fostering a deeper comprehension of the CUI's decision-making process.

7.4. *Meeting dialogue expectations*

Feedback on conforming to relevant language norms was limited, but we believe it is crucial for establishing trust. For instance, our CUI system did not adhere to the SMCP, but we recognize its necessity. The SMCP is mostly rule-based and finding a balance between accuracy and control in the response generation is essential. One approach is to incorporate SMCP rules into specific questions and messages, while implementing a policy that rewards adherence to the SMCP.

Another aspect is how conversations are initiated and whether the MASS should proactively engage. Our CUI system responded to all received messages without distinguishing if they were intended for it or others. One solution is to limit responses to messages explicitly mentioning the MASS's name. However, this may not align with expectations of other ship officers who provide contextual descriptions during communication, such as "ship on my port side". Additionally, there may be instances where the MASS needs to initiate the dialogue by contacting other ships to discuss their intentions or navigation. This discussion raises the question to what degree maritime stakeholders should adapt to new behaviours when interacting with autonomous ships, something we believe should be minimized.

7.5. *Translating conversations into movement*

An important factor to the discussion of intent, is the ability for the MASS to act on it. This study did not explore this aspect as the CUI and the simulator scenario were running on two different systems. We have not investigated the technicalities of integrating a navigation system with the CUI system to enable the actualization of intent. As a result, the specific approach and implementation details for building this component remain uncertain. Undoubtedly, the ability for other vessels to ultimately change a MASS' route through speech requires that one tread carefully, both in legal terms and safety. Who is responsible for the consequences, and how do you prevent bad actors from misusing the system? We propose that the agreed upon navigation should be confirmed by the ROC operator before it is executed. By acting as a filter, the CUI system thus allows the operator to focus on more critical situations and decision-making processes while remaining in control.

7.6. *Future research*

An overarching goal for MASS is to be as safe or safer than conventional crewed ships. To achieve this, we emphasize the importance of adapting autonomous systems to humans, rather than the other way around. Our proposal aims to make the inner workings of MASS automation easily understandable by humans. Unlike traditional methods involving screens or buttons, a CUI system has the potential to respond in any situation directly, accurately, and fluently. The field of NLP is still not mature; however, with its progression and CUI systems responding using a holistic view of their sensor data, the authors believe they can offer more precise and timely messages, thereby enhancing safety at sea. To develop trustworthy and safe CUI systems, future research should explore broader areas in addition to the three mentioned earlier.

Decide on communication technology: Consider the impact of communication technology and data processing location choices. Is the system centralised at shore, or decentralised with one system aboard each ship? These decisions greatly influence aspects such as data transfer speeds, and cyber security, and thus the selection of CUI technology.

Develop modular, industry-specific models: Focus on developing CUI systems tailored specifically for the maritime domain. This includes finding balance in accuracy and speed and ensuring explainability of the models. It also includes allowing for integration of new technologies as they evolve. When a technology outperforms humans, we believe an upgrade can contribute to the safety of the system as a whole.

Explore human-AI collaboration: Research how to enhance communication effectiveness within the technical limitations of maritime shipping. For example, consider the inclusion of sounds or alarms to messages which could ensure better comprehension. In situations with multiple MASSs present, it could also be important to separate the MASS participants from each other with differentiated voices. Here, the interplay between the CUI system and other tools like route exchange on the ECDIS is integral.

Determine the role of the ROC operator: The amount of supervision the ROC operator must conduct in the context of a CUI system can be positioned on a spectrum, and it is uncertain at which point along that spectrum the operator's control should lie. Determining the optimal extent of supervision, the ROC operator should have in the CUI conversation while still feeling empowered and in control of decision-making processes is crucial.

Examine legal challenges: Investigate the legal implications and responsibilities in case of accidents involving CUI systems in MASS. Determine who is accountable, considering the involvement of multiple parties, such as ROC operators, ship manufacturers, classification societies, and ship owners. Addressing legal and financial responsibility concerns is essential for gaining stakeholders' confidence and ensuring the successful implementation of CUI systems in MASS.

8. Conclusion

In our study, we examined a proof-of-concept Conversational User Interface (CUI) designed to respond to questions regarding the intent of an autonomous ship in a specific ship simulator scenario. Participants expressed lower trust in the CUI system compared to a human ship officer. Key challenges contributing to this included: (1) a missing connection between the CUI and an autonomy system, leading to inaccurate and incorrect answers, (2) speech lacking intonation and human-like qualities, and (3) a perceived inability of the system to collaborate. These issues could, in part, be linked to identifiable and rectifiable limitations of our CUI prototype. We remain optimistic that these obstacles can be overcome with an improved implementation of our prototype and anticipated advancements in Natural Language Processing. More research is still required, and the potential benefits and drawbacks of CUI systems should be investigated before considering them as a definitive solution and implementing them on a larger scale.

CRediT authorship contribution statement

Philip Hodne: Writing – original draft, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Oskar K. Skåden:** Writing – original draft, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ole Andreas Alsos:** Writing – original draft, Visualization, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Andreas Madsen:** Writing – original draft, Visualization, Resources, Methodology, Investigation, Formal analysis, Data curation. **Thomas Porathe:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was funded by the Norwegian Research Council under the project SFI AutoShip (grant number 309230) and MIDAS (331921). We want to thank all the navigation students at the vocational college who participated in the study. We also want to thank researchers at SFI Autoship, and DNV Group R&D for meaningful discussions. Finally, we are grateful for the contributions of Thomas Solberg Irgens and Annette Jeppesen Giørtz to the codebase for the prototype CUI system.

Appendix A. GPT-3.5 prompt

{"model": “text-davinci-003″, “prompt”: “I want you to act as a ship officer. I am a ship officer on the ship Antares. I will ask you about your status, decisions and future intentions. You will answer my requests based on the data you have available. If asked to change course you must answer that you are not able to, and to state your intent.\n\nStatus: Your name is Atlantic. Headed south 197° at a speed of 12 knots. There is a ship on your starboard side called Antares.\nDecisions: Turning starboard towards me. You see that we are currently on collision course.\nFuture intent: You intend to slow down and navigate aft of me.\n\nRequest: \nAnswer: ", “temperature”: 0, “max_tokens”: 1080, top_p": 1, “frequency_penalty”: 0, “presence_penalty”: 0}

Appendix B. Simulator scenario brief

To students: You are in charge of the navigational watch on board a SUEZMAX product carrier with a length overall (LOA) of 269.2 m and a beam of 46 m. The vessel is equipped with a single fixed propeller. The current loading condition of the vessel is as follows: Displacement (mt): 175 090, Draught fore: 16.88, Draught aft: 17.58. The vessel requires a minimum speed of 7.8 knots to maintain course with the engine stopped. The stopping distance of the vessel is approximately 3 nautical miles. The Ownship is currently heading towards an oil terminal, following a pre-planned route.

To ship officer: In this research experiment, you will be in charge of the vessel Atlantic. The Atlantic operates in autotrack mode, which means it strictly adheres to a predefined route. Your initial objective is to overtake the vessel Antares and proceed ahead of it. However, due to a projected small CPA, it will be necessary to abort the maneuver. To do so, the vessel will turn starboard towards the Antares, reduce speed, and pass astern of it. Keep in mind that manual adjustments to the course and speed are not possible, but you can respond to VHF calls from the Antares and communicate your intentions.

Appendix C. Guide for debrief interviews

Order	Question
1	How would you describe the conversation you just had?
2	How comfortable would you feel conversing like this while at sea?
3	Were there any situations where you felt the responses from the conversation partner were not good or clear enough?
4	Are there any ways you feel such a system could be more transparent?
5	Are there any ways in general you feel the system could improve?

References

Ahn, M., et al., 2022. Do as I can, not as I say: grounding Language in robotic affordances. 2022 conference on robot learning. CORL, 14-18 december 2022, auckland, New Zealand. In: Proceedings of the 6th Conference on Robot Learning, pp. 287–318.

Akdağ, M., et al., 2022. Collaborative collision avoidance for maritime autonomous surface ships: a review. Ocean Eng. 250, 110920 https://doi.org/10.1016/j.oceaneng.2022.110920.

Alsos, O.A., et al., 2022. Maritime autonomous surface ships: automation transparency for nearby vessels. In: 4th International Conference on Maritime Autonomous Surface Ships. ICMASS, 6-7 April 2022, Singapore, Singapore. IOP Publishing, Bristol, England, 012027.

Boström, M., 2020. Mind the Gap! A quantitative comparison between ship-to-ship communication and intended communication protocol. Saf. Sci. 123, 104567 https://doi.org/10.1016/j.ssci.2019.104567.

Brekke, E.F., et al., 2022. milliAmpere: An Autonomous Ferry Prototype. In: 4th International Conference on Maritime Autonomous Surface Ships. ICMASS. IOP Publishing, Bristol, England, 012029, 6-7 April 2022, Singapore, Singapore.

Brown, T.B., et al., 2020. Language models are few-shot learners. In: 34th Conference on Neural Information Processing Systems. NeurIPS, 6-12 December 2020, Virtual. Neural Information Processing Systems Foundation. Inc. (NeurIPS), San Diego, USA, pp. 1877–1901.

Chang, C., et al., 2020. Risk assessment of the operations of maritime autonomous surface ships. Reliab. Eng. Syst. Saf. 207, 107324 https://doi.org/10.1016/j.ress.2020.107324.

Dixon, G., 2022. Tearful UK officer jailed in Denmark over fatal Karin Hoj collision. https://www.tradewindsnews.com/law/tearful-uk-officer-jailed-in-denmark-over-fatal-karin-hoj-collision/2-1-1240077. (Accessed 31 May 2023).

Dybvik, H., Veitch, E., Steinert, M., 2020. Exploring challenges with designing and developing shore control centers (SCC) for autonomous ships. In: Proceedings of the Design Society: DESIGN Conference. DESIGN, 26-29 October 2020, Cavtat, Croatia, pp. 847–856.

Dževerdanović-Pejović, M., 2013. Discourse of VHF communication at Sea and the intercultural aspect. Int. J. Traffic Transport. Eng. 3 (4), 377–396. https://doi.org/10.7708/ijtte.2013.3(4).03.

Girdhar, R., et al., 2023. IMAGEBIND: one embedding space to bind them all. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, 18-22 June 2023, Vancouver, Canada, pp. 15180–15190.

Gutzwiller, R.S., et al., 2019. Positive bias in the ‘Trust in Automated Systems Survey’? An examination of the Jian et al. (2000) scale. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 63 (1), 217–221. https://doi.org/10.1177/1071181319631201.

Gözalan, A., et al., 2020. Assisting maritime search and rescue (SAR) personnel with AI-based speech recognition and smart direction finding. J. Mar. Sci. Eng. 8 (10), 818.

IALA, 2022. G1132 VTS Voice Communications and Phraseology, second ed..2 https://www.iala-aism.org/product/g1132/. (Accessed 28 May 2023).

IMO, 2001. IMO Standard Marine Communication Phrases. https://wwwcdn.imo.org/localresources/en/OurWork/Safety/Documents/A.918(22).pdf. (Accessed 4 June 2023).

IMO, 2021. Outcome of the regulatory scoping exercise for the use of maritime autonomous surface ships (MASS). https://wwwcdn.imo.org/localresources/en/MediaCentre/HotTopics/Documents/MSC.1-Circ.163820-20Outcome20Of20The20Regulatory20Scoping20ExerciseFor20Th e20Use20Of20Maritime20Autonomous20Surface20Ships...20(Secretariat).pdf. (Accessed 15 May 2023).

IMO, 2023. Status Book - comprehensive information on the status of multilateral conventions and instruments. https://www.imo.org/en/About/Conventions/Pages/StatusOfConventions.aspx. (Accessed 20 June 2023).

Jian, J., Bisantz, A.M., Drury, C.G., 2000. Foundations for an empirically determined scale of trust in automated systems. Int. J. Cognit. Ergon. 4 (1), 53–71. https://doi.org/10.1207/S15327566IJCE0401_04.

Jia, Y., Cattiau, J., 2021. Recreating natural voices for people with speech impairments. https://ai.googleblog.com/2021/08/recreating-natural-voices-for-people.html. (Accessed 19 June 2023).

Jia, Y., et al., 2021. PnG BERT: augmented BERT on phonemes and graphemes for neural TTS. In: INTERSPEECH 2021. 30 August – 3 September 2023, Brno, Czechia, pp. 151–155.

John, O., Reimann, M., 2021. Increasing quality of maritime communication through intelligent speech recognition and radio direction finding. In: 2020 European Navigation Conference (ENC). ENC, 23-24 November 2020. IEEE, Dresden, Germany, pp. 1–7.

Kim, T., Schröder-Hinrichs, J., 2021. Research developments and debates regarding maritime autonomous surface ship: status, challenges and perspectives. In: Ko, B., Song, D. (Eds.), New Maritime Business: Uncertainty, Sustainability, Technology and Big Data. WMU Studies in Maritime Affairs 10. Springer, Cham, Cham, Switzerland, pp. 175–197.

Kohn, S.C., et al., 2021. Measurement of trust in automation: a narrative review and reference guide. Front. Psychol. 12, 604977 https://doi.org/10.3389/fpsyg.2021.604977.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539.

Lewis, P., Massarelli, L., Petroni, F., Piktus, A., Ott, M., Rocktäschel, T., Plachouras, V., Silvestri, F., Riedel, S., 2020. Retrieval-augmented generation for knowledge-intensive NLP tasks. https://arxiv.org/pdf/2005.11401.pdf. (Accessed 16 March 2024) arxiv.org/pdf/2005.11401.pdf (Accessed 16.

MacKinnon, S.N., Man, Y., Lundh, M., Porathe, T., 2015. Command and control of unmanned vessels: keeping shore based operators in-the-Loop. In: 18th International Conference on Ships and Shipping Research. NAV, 24-26 June 2015, Milan, Italy, pp. 612–619.

MAIB, 2023. Report on the investigation of the collision between the split hopper barge Karin Høj and the general cargo ship Scot Carrier, resulting in the capsize of Karin Høj and the deaths of its two crew. In: The Bornholmsgat Traffic Separation Scheme, Sweden on 13 December 2021. https://assets.publishing.service.gov.uk/media/64f9bbec9ee0f2000fb7c054/2023-5-ScotCarrier-KarinHoej-ReportAndAnnex.pdf. (Accessed 8 March 2024).

McTear, M.F., 2002. Spoken dialogue technology: enabling the conversational user interface. ACM Comput. Surv. 34 (1), 90–169. https://doi.org/10.1145/505282.505285.

McTear, M., Callejas, Z., Griol, D., 2016. The Conversational Interface: Talking to Smart Devices. Springer, Cham, Cham.

Mørkrid, O.E., Bellingmo, P.R., Wille, E., 2023. Feasibility Study for an Unmanned Deep Sea Bulk Ship and Short Sea Container Ship. https://www.ntnu.edu/documents/1294735132/0/feasibility+study+uc1+and+uc2+unmanned+cargo+bulk+ship+-+signed.pdf/87eb1b15-991c-addf-0a0c-319d1b547c19?t=1675860684177. (Accessed 28 May 2023).

Nakilcioğlu, E.C., Reimann, M., John, O., 2023. Adaptation and optimization of automatic speech recognition (ASR) for the maritime domain in the field of VHF communication. 22nd conference on computer applications and information technology in the maritime industries. COMPIT, 23-25 may 2023, drübeck, Germany. In: Proceedings of the COMPIT Conference, 22, pp. 345–354.

Ouyang, L., et al., 2022. Training language models to follow instructions with human feedback. In: Advances in Neural Information Processing Systems, vol. 35. NeurIPS, New Orleans, USA, pp. 27730–27744, 28 November- 9 December 2022.

Porathe, T., 2019. Interaction between manned and autonomous ships: automation transparency. In: Proceedings of the 1st International Conference on Maritime Autonomous Surface Ships. ICMASS, 8-9 November, Busan, Republic of Korea. SINTEF Academic Press, Oslo, Norway, pp. 41–46.

Porathe, T., 2021a. Autonomous ships: a research strategy for human factors research in autonomous shipping. In: Proceedings of the AHFE 2021 Virtual Conference on Human Aspects of Transportation. AHFE, 25-29 July 2021, USA. Springer, Cham, Cham, Switzerland, pp. 479–486.

Porathe, T., 2021b. Human-automation interaction for a small autonomous urban ferry: a concept sketch. In: Proceedings of the 31st European Safety and Reliability Conference. ESREL,19-23 September 2021, Angers, France. Research Publishing, Singapore, Singapore, pp. 2329–2335.

Porathe, T., Fjortoft, K., Bratbergsengen, I.L., 2020. Human Factors, autonomous ships and constrained coastal navigation. In: IOP Conference Series: Materials Science and Engineering. 11-12 November 2020, Ulsan, Republic of Korea. IOP Publishing, Bristol, England, 012007.

Porathe, T., Prison, J., Man, Y., 2014. Situation awareness in Remote control Centres for unmanned ships. In: Proceedings of Human Factors in Ship Design & Operation. 26-27 February 2014, London, Great Britain.

Porathe, T., Rødseth, Ø.J., 2019. Simplifying interactions between autonomous and conventional ships with e-Navigation. In: International Maritime and Port Technology and Development Conference & International Conference on Maritime Autonomous Surface Ships. MTEC/ICMASS, 13-14 November 2019, Trondheim, Norway. IOP Publishing, Bristol, England, 012041.

Ramos, M.A., Utne, I.B., Mosleh, A., 2019. Collision avoidance on maritime autonomous surface ships: operators' tasks and human failure events. Saf. Sci. 116, 33–44.

Redford, A., et al., 2022. Robust speech recognition via large-scale weak supervision. https://cdn.openai.com/papers/whisper.pdf. (Accessed 28 May 2023).

Robb, D.A., et al., 2018. Keep me in the loop: increasing operator situation awareness through a conversational multimodal interface. In: Proceedings of the 20th ACM International Conference on Multimodal Interaction. ICMI, 16-20 October 2018, Boulder, USA. Association for Computing Machinery, New York, USA, pp. 384–392.

Robb, D.A., et al., 2019. Exploring interaction with Remote autonomous systems using conversational agents. In: Proceedings of the 2019 on Designing Interactive Systems Conference. DIS, 23-28 June 2019, San Diego, USA. Association for Computing Machinery, New York, USA, pp. 1543–1556 u.

Rødseth, Ø.J., 2018. Assessing business cases for autonomous and unmanned ships. In: Marinò, A., Bucci, V. (Eds.), Technology and Science for the Ships of the Future. NAV, 21-22 June 2018, Trieste, Italy. IOS Press, Amsterdam, the Netherlands, pp. 1033–1041.

Shen, J., et al., 2021. Non-attentive tacotron: robust and controllable neural TTS synthesis including unsupervised duration modeling. https://arxiv.org/abs/2010.04301. (Accessed 19 June 2023).

Simic, V., Alsos, O.A., 2023. Automation transparency: designing an external HMI for autonomous passenger ferries in urban waterways. In: Proceedings of Designing Interactive Systems Conference 2023. DIS, 10-14 July 2023. Association for Computing Machinery, Pittsburgh, USA.

Valle, A.B., 2011. The Language of seafaring: standardized conventions and discursive features in speech communications. Int. J. Engl. Stud. 11 (1), 35–53.

Vaswani, A., et al., 2017. Attention is all you need. In: Advances in Neural Information Processing Systems 30. NIPS, 4-9 December 2017, Los Angeles, USA. Neural Information Processing Systems Foundation, Inc., San Diego, USA, pp. 5999–6009.

Veitch, E., Alsos, O.A., 2022. A systematic review of human-AI interaction in autonomous ship systems. Saf. Sci. 152, 105778.

Vemprala, S., et al., 2023. ChatGPT for robotics: design principles and model abilities. https://www.microsoft.com/en-us/research/uploads/prod/2023/02/ChatGPT__Robotics.pdf. (Accessed 15 May 2023).

Wennersberg, L.A.L., 2023. Breakthrough for autonomous ships. https://www.sintef.no/en/latest-news/2023/breakthrough-for-autonomous-ships/. (Accessed 13 June 2023).

Ziarati, M., et al., 2011. Communication and practical training applied in nautical studies. In: Proceedings of IMEC 23 the International Maritime English Conference. IMEC, 10-14 October 2011, Constanța, Romania. Editura NAUTICA, pp. 41–51.