Journal of Physics: Conference Series

PAPER • OPEN ACCESS

On the use of maritime training simulators with humans in the loop for understanding and evaluating algorithms for autonomous vessels

To cite this article: Anete Vagale *et al* 2022 *J. Phys.: Conf. Ser.* **2311** 012026

On the use of maritime training simulators with humans in the loop for understanding and evaluating algorithms for autonomous vessels

Anete Vagale[1], Ottar L Osen[1,2], Andreas Brandsæter[1], Marius Tannum[2], Christian Hovden[2] and Robin T Bye[1]

[1] Cyber-Physical Systems Laboratory, Department of ICT and Natural Sciences, NTNU – Norwegian University of Science and Technology, Postboks 1517, NO-6025 Ålesund, Norway
[2] Faculty of Technology, Natural Sciences and Maritime Sciences, University of South-Eastern Norway, Postboks 235, NO-3603 Kongsberg, Norway

E-mail: anete.vagale@ntnu.no

Abstract. The prospect of a future where the maritime shipping industry is dominated by autonomous vessels is appealing and gaining global interest from industry majors, research institutions, and academia. Potential advantages include increased operational safety, reduced costs, and lower environmental footprint. However, the transition will not happen overnight and is not without challenges. For example, algorithms for autonomous navigation must take into consideration safety concerns of the own ship, its crew and passengers, other surrounding ships, and the surrounding environment. This raises a need to test and verify safety, performance, and robustness of the algorithms responsible for the autonomous functionality. In addition, the transition towards fully autonomous ships is likely to be gradual and involve remote control centres and ships with varying degrees of autonomy. Hence, humans will inevitably have to interact with autonomous vessels in a variety of scenarios, including overriding own ships from land or on board, as well as communicating with autonomous ships from other fleets. Inevitably, full scale scenario testing involving real vessels and humans is costly, impractical, time-consuming, and potentially dangerous. In this paper, we propose an alternative approach, and explore how maritime navigation training simulators with humans in the loop can be used as a testbed for understanding and evaluating algorithms for autonomous vessels. In the proposed setting, we can directly compare choices made by an algorithm with those of a skilled human navigator for a variety of navigational tasks. Moreover, we can study in real-time the behaviour and decision-making of human navigators in mixed scenarios that also include autonomous ships, whether this is known beforehand or not. Our paper provides an overview of related work, details on maritime simulators and how algorithms can be tested, and some of the technical requirements. To exemplify our approach, we present two example test setups, and provide a brief discussion of our findings. We conclude that using maritime training simulators enables the study of several interesting and vital research questions, including that of the interaction between autonomous and traditional vessels operating side by side.

1. Introduction

With autonomous ships on the horizon that may affect the safety of the own ship and other ships, its crew and passengers, and the surrounding environment, the need to test and verify the safety, performance and robustness of the autonomous functionality is a top priority. Autonomous ships therefore require sophisticated sensors and algorithms to collect and fuse information about the ship and its surroundings to form an adequate situational awareness. Based on the situational awareness, an advanced navigation algorithm[1] must be able to automatically change the ship's route and perform collision avoidance manoeuvres, ensuring safe sailing in compliance with rules and regulations. Lack of unambiguous specifications is also a challenge for an autonomous ship, since current regulations, including The Convention on the International Regulations for Preventing Collisions at Sea (COLREG) [1], use expressions which require interpretations, such as ample time, good seamanship, safe speed, etc. For an overview of relevant intelligent autonomous navigation algorithms, see [2].

Due to the complexity of navigating a vessel, traditional algorithms from control engineering and other classic fields can easily come up short, with algorithms from the field of artificial intelligence (AI) such as deep neural networks, evolutionary algorithms, and reinforcement learning tending to be better choices. However, AI-based algorithms and particularly machine learning algorithms often suffer from a lack of explainability or interpretability, and they may also be inherently complex. As a consequence, testing such algorithms quickly becomes a complex problem in itself. Full scale realistic testing of algorithms employed on a real ship is inevitably going to be costly, dangerous, impractical and probably not even allowed without thorough testing in a safe environment in advance. Two existing testing approaches are: (i) simulation-based testing and (ii) functionality-based testing. In this paper, we discuss and focus on the simulation-based testing approach. A procedure for simulation-based verification of autonomous navigation systems is proposed in [3]. The verification and assurance of machine learning algorithms needs to be fundamentally different from traditional assurance and verification processes based on requirements and physical understanding [4, 5, 6, 7]. Moreover, since autonomous vessels will have to relate to traditional vessels with humans onboard and vice versa, the human dimension becomes very important.

There are several ship simulators, e.g., [8, 9, 10], that could enable us to test path planning and collision avoidance scenarios but as described later in this article, there are many challenges that require more advanced setups. For example, in order to evaluate interactions with humans, there is a need for a very realistic setup. Likewise, algorithms could be tested by simulating data from the automatic identification system (AIS) during realistic navigation scenarios but the lack of a visual model for the AIS-simulated vessel will impact the human navigator's decision-making. It will therefore be of limited value to study the human navigator's actions towards an invisible AIS-simulated "ghost ship."

Due to humans' ability to learn and adapt to new and unexpected situations, humans can play an important role in complex technological systems such as autonomous navigation to ensure safe and efficient operation [11]. Testing how human navigators will react to an autonomous vessel piloted by an algorithm in real life is one of several interesting research questions. It could also be valuable to compare how well an algorithm solves a given set of scenarios compared to a human navigator.

In this article, we argue that the use of maritime training simulators will enable testing and research that is otherwise very difficult to perform, and propose to use these simulators as a testbed to investigate research questions similar to the above. The proposed test setups could aid in evaluating how well algorithms for autonomous vessels perform in general and especially with humans in the loop. We suggest looking both at a setup when an intelligent algorithm is controlling target ship(s) in the simulator, and a setup where an intelligent algorithm is controlling the bridge and executing navigation tasks. Another benefit of using a simulator for these experiments is the ability to control the experiment and making sure that the situations can be replicated.

[1] In this paper, the word 'algorithm' may also refer to several algorithms combined and used together, or an autonomous system as a whole.

The remainder of the paper is organized as follows: Related work is reviewed in Section 2. Section 3 provides an overview of existing maritime simulators, their history and possible simulator setups. Testing of autonomous algorithms is described in Section 4. Section 5 presents technical requirements for the proposed use cases. Different test setups are proposed in Section 6. Section 7 contains a discussion, and finally, some concluding remarks are drawn in Section 8.

2. Related work

Autonomous vessels are already investigated using various simulators. However, using *maritime training simulators* normally used in education of cadets, as proposed in this paper, is a novelty. Performing tests in maritime training simulators enables a means for verifying autonomous navigation systems or path planning algorithms, with humans in the loop, in a safe simulated environment before implementing them on a real vessel at sea.

A maritime training simulator that could fulfill this purpose is the K-Sim simulation platform developed by Kongsberg Digital, certified by DNV and fulfilling Standards of Training, Certification and Watchkeeping (STCW) requirements. Its intended purpose relates to virtual prototyping, testing, and verification of autonomous algorithms for vessels [12]. This includes scenario generation faster than real-time, energy consumption predictions, taking into account environmental conditions, and route verifications. Another use case of the platform is simulation-based prediction, as well as using live data from a ship and its environment.

In another project, Wärtsilä delivered a navigation simulator and specific mathematical models for Intelligent Shipping Technology Test Laboratory (ISTLAB) at the Satakunta University of Applied Sciences (SAMK) in Finland [13]. The goal of that project, launched in early 2019, was to develop a testing environment for remotely controlled, autonomous vessels.

With a focus on simulation-based verification of autonomous navigation systems, Pedersen et al. [3] propose to use a digital twin for testing purposes. A complete test system includes scenario manager, test evaluation module, operating environment, test interface, and digital twin that represents the systems of the own ship. The authors indicate that the Open Simulation Platform (OSP) [14] may potentially be used later for testing purposes.

Finally, we draw attention to research by Vagale et al. [15], who focus specifically on the evaluation of path planning algorithms from the risk and safety perspective.

3. Maritime training simulators

Maritime simulators for training can perhaps be labelled ”very serious games.” These simulators have a very realistic touch and feel since they typically have a vessel bridge replica with large screens with a wide view angle, all essential handles and equipment found on board a real ship, and often even an audio system that provides realistic sounds. This very realistic physical setup combined with strict rules on expected behaviour, and maybe even combined with use of uniforms, makes it possible to create a theatre-like atmosphere that is immersive and blur the lines between simulator and real life. Indeed, the provided training in these simulators is considered realistic enough that, according to International Maritime Organization [16], cadets are permitted to fulfil part of their training in a simulator.

Although marine training simulators may seem realistic to human navigators, it cannot be assumed that the simulators are sufficiently realistic for testing a particular class of autonomous vessel functionality, namely that of object detection and classification. Hence, this functionality cannot be directly implemented in the simulators. To overcome this challenge, authors in [17] propose utilising Cycle-Consistent Adversarial Networks (Cycle-GANs) to transfer the simulator data to a real-world-like environment before autonomous functionality such as object detection and classification is performed.

3.1. History

Technically sophisticated simulators are in particular found in the training of professionals where one seeks to increase safety, such as aviation, shipping, and healthcare [18]. The earliest ship bridge simulators with visual night scenes were introduced in the 1970s [19] and full bridge simulators have already been commercially available for decades, reaching the point where the international rules and regulations as prescribed by the International Maritime Organization (IMO) mandates the use of simulator training as a basis for certifying professional navigators through its STCW guidelines.

The 2010 amendments of the STCW 1978 code section A-I/12 added requirements for its member states to include simulator-based training with modern technology like electronic chart displays information system (ECDIS), AIS, and dynamic positioning (DP). The updated codes also define that seafarers may demonstrate their competence to handle a ship in all conditions through approved simulator training while at the same time providing the nautical student with limited credit as an equivalency for sea-going service depending on factors comprising the level of simulation and scenarios, time spent, student-teacher ratio, pre-brief and de-brief procedures, and integration with other elements in the approved training program [16]. The STCW requires the approval of simulators used for mandatory training or assessment of seafarers, and classification societies often do the travail of simulation certification. Det Norske Veritas (DNV) for instance has developed a standard for certifying simulator compliance, where they also define four classes of bridge operation simulators, Class A, B, C and S, where only Class A, full mission simulator, meets the IMO requirement of "manoeuvre and handle a ship in all conditions" demanded for approval of sea-time equivalency [20].

3.2. Major vendors

Major vendors of full mission maritime training simulators in accordance with STCW 2010 are:

- K-SIM by Kongsberg Digital [8]
- NTPRO 5000 by Wärtsilä [9]
- NAUTIS by VSTEP [10]
- SIMFLEX by FORCE Technology [21]
- BOREALIS by Poseidon [22]
- REMBRANDT by BMT [23]
- ARI SMS by ARI Simulations [24]
- IS FMBS by Image Soft [25]
- ANS6000 by Rheinmetall Electronics [26]

3.3. Simulator setup

Typically, a simulator setup consists of at least two or more rooms (one-to-many setup). Firstly a room where the main instructor controls the exercise, configures the training scenario, selects vessel types and sizes, adjusts weather parameters, and more. From this room, the instructor may monitor and record the cadets' performance during the exercise. In simpler cases, a one-to-one simulator with only one navigation bridge connected to the instructor station is also possible (see Fig. 1). An example setup of five navigation bridges connected to a main instructor station is given in Fig. 2.

In addition, the simulator can contain one or more vessel bridge replicas. For simple exercises a single bridge is sufficient, however, both in order to train on more complex challenges and to train more cadets at the same time it is convenient to use several bridges. Typically, the bridges are linked with the same simulation environment, thus allowing the different bridges (simulated vessels) to interact with each other. A picture of a typical training bridge is shown in Fig. 3.

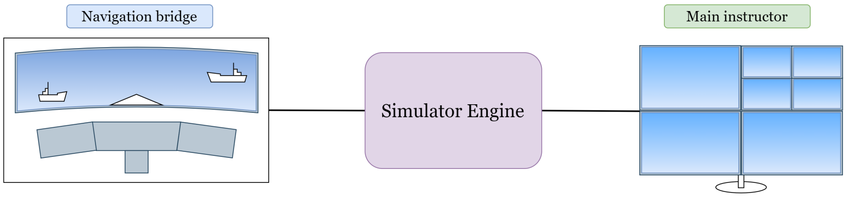

Figure 1. Setup of a one-to-one navigation training bridge connected to the main instructor station through a simulator engine.

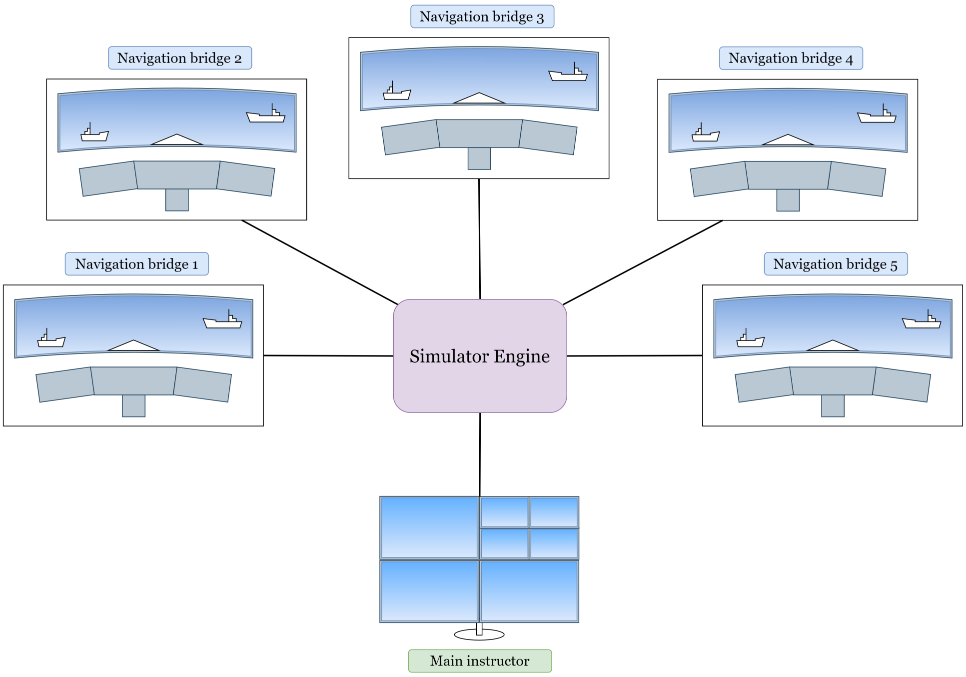

Figure 2. One-to-many setup of the main instructor station (in the centre) connected to five navigation training bridges through a simulator engine.

An example of an instructor station overlooking the actions of the cadet navigating the bridge is shown in Fig. 4. A research training simulator is shown in Fig. 5.

3.4. Target vessels
In addition to one or more training bridges, the operator or instructor has the possibility to control a number of additional vessels (target vessels) in the training scenario. These vessels have no separate simulator bridge, but are merely vessels, chosen from a library of vessels, that are placed into the simulation environment and given a heading and speed. Throughout the exercise, the operator may adjust these vessels' heading and speed. Later in this paper, we present setups where the target vessels are controlled by autonomous path planning and collision avoidance algorithms.

Figure 3. Navigation training bridge at the Department of the Ocean Operations and Civil Engineering at NTNU in Ålesund. Photo: Anete Vagale.

Figure 4. Instructor station at the Department of the Ocean Operations and Civil Engineering at NTNU in Ålesund. Photo: Terje Ole Slinning.

Figure 5. Research training bridge at the Department of the Ocean Operations and Civil Engineering at NTNU in Ålesund. The research bridge consists of the navigation bridge with controls, three screens (on the right), and the instructor's area (on the left). Photo: Anete Vagale.

4. Testing of algorithms

In order to evaluate the performance and capacities of algorithms, extensive testing is required. These tests should be standardized, reproducible, relevant and realistic. As of today, we are not aware of any existing standardized tests for autonomous ships. An attempt at defining requirements for simulator based testing is given in [27]. We expect different actors to develop their own tests before some governing bodies establish some official tests, especially related to safety. However, in addition, there will probably be several other "industry standards" that relate to issues such as efficiency.

4.1. Challenges with assurance of machine learning algorithms

Traditional verification, including the V-model (e.g., see ISO 26262 *Road vehicles - functional safety* in [28]), typically assumes that the requirements of a component are completely specified and that "each refinement can be verified with respect to its specification" [29]. Problems like machine perception and ship navigation cannot be clearly specified. For example, COLREGs [1] that must be complied with when navigating on waters, are in many cases open for interpretation. This has motivated the use of machine learning algorithms which learn from examples rather than being programmed based on a specification [5].

However, understanding and interpreting a machine learning algorithm's reasoning is challenging or even impossible, which makes it challenging to verify the algorithm or to determine when the algorithm's reasoning is in error [30, 31, 32]. Explainable AI (XAI) for algorithms attempts to give a reasoning for the decisions suggested by an algorithm. Additionally, machine learning algorithms are completely dependent on the quality of its training data. Hence, any verification and assurance process needs to include a comprehensive data analysis, documenting if the dataset sufficiently covers the input domain, is representative and complete, particularly regarding corner cases [4]. The choice of resampling strategy, that is, how the data is split into training and test sets using cross validation or bootstrap techniques, must also be evaluated.

The robustness of the algorithms must also be tested and documented. Brandsæter at el. [33] provide an example showing how minor, seemingly insignificant changes like a small image rotation can confuse a classifier. In their example, a one-degree rotation causes a misdetection of a vessel. It is also shown that when the image is rotated further (three degrees), the algorithm once again correctly detects the vessel, illustrating the unpredictable behaviour of the algorithm. Image manipulations and transformations utilising, for example, Generative Adversarial Networks (GAN) (e.g., [34, 35]) to transform a scene from summer to winter or daylight to night, as well as simple augmentations including rotations, share, blur, and similar (e.g., see [36, 37, 7]) can be utilised to increase the test scope of a limited dataset.

4.2. Metrics

The capability and performance of detecting other ships is an essential part of an autonomous ship's situational awareness capability. Relevant metrics for object detection includes number of correctly detected ships (true positives), but also missed targets (false negatives), as well as true negatives and false positives. These numbers are often summarised in other metrics such as precision, recall / sensitivity (also known as true positive rate), specificity (also called true negative rate), and F1-score. But are all targets equally important? Detecting a ship which is far away is less important than detecting a nearby ship on collision course. As the relevance of the different targets depends on the navigation, this should be taken into consideration. Hence, assessing the ship's situational awareness in isolation is not recommended, although methodology for assessing the situational awareness of humans is available, e.g., see [38, 39, 40].

Øvergård et al. [41] argue that existing research is suggestive of limitations of the reliability of subjective assessments, and propose an initial version (prototype) of an automated assessment algorithm for a specific maritime operation. The following control requirements for the safety of navigation, collected based on open interviews with six subject-matter experts who all held deck-officer certifications, were used in the prototype:

(i) distance to land based on own ship length,
(ii) distance to moving objects (vessels) based on own ship length,
(iii) distance to floating objects based on ship length,
(iv) the deviation between ship heading and heading of dock (meaning that the ship should be parallel to the dock during the last part of docking), and
(v) the minimum depth below the ship's keel (the so-called 'safety depth').

The aforementioned parameters can be considered a part of a greater geometric collision risk assessment (CRA) set. More information on metrics for CRA for autonomous vessels, and a comparison of 45 different path planning and collision algorithms, is given in [42].

On the other hand, Vagale et al. [15] propose introducing additional relevant metrics to evaluate path planning algorithms of autonomous surface vehicles (ASVs), such as:

- efficiency / path fitness (e.g. length of path / time used / fuel used),
- compliance with regulations (COLREGs), and
- good seamanship practice.

The above-mentioned metrics are still not clearly defined and in some cases could be overlapping. Typically, compliance with regulations and safety often overlap. Similarly, good seamanship practice requires following COLREGs and vice versa. It is important to look at these metrics in connection with each other; the safest move might be not to move at all, however, that would obviously affect efficiency negatively.

4.3. Turing test for autonomous ships

In addition to evaluating the above-mentioned metrics, interviewing cadets and instructors after the exercise might be useful to capture their experience with the autonomous ship. In this case, it could also be interesting to consider doing blind tests where the cadets and/or instructors do not know which ships are autonomous (similar to Turing's imitation game).

Usually, the cadets train on scenarios where the other vessels (target vessels) are navigated by the instructor. In such a setup, some (or all) target vessels could be autonomous ships. It would then be interesting to see if the cadets notice any difference. An even more interesting scenario emerges from linking several simulator bridges together (see Figure 2) to share a common scene. For a cadet, any observed ship could either be a target ship placed there by the instructor, another ship operated by another cadet(s), or in our case, an autonomous ship. It would be impossible to know without interacting or studying the ship's behaviour.

Finally, it could also be very interesting to do such an experiment in a way that the instructor played the Turing game. Either by letting him / her navigate a simulator bridge instead of the cadet(s), or by manipulating the instructor panel in such a way that there is no way of knowing which ship is autonomous and which is manually steered. This experiment is of great interest because the instructor have years of experience monitoring cadets' navigational behaviour.

How could an autonomous vessel divulge its nature? For one, by performing actions that are contradictory to good seamanship and/or COLREG rules. For instance, several algorithms that have been proposed, e.g., [43, 44], would change heading very often or continuously. This is against COLREG rule 8 since your intentions should be communicated clearly and your manoeuvres should be predictable. This is typically done by navigating in straight lines and making as few heading changes (way-points) as possible. This kind of navigation will often be in stark contrast to many algorithms that focus on optimization and rely on cost or fitness functions. Another way of identifying vessels is radio communication. This is an integral and important part of the cadet's training. Training without the use of radio would usually be seen as out of the ordinary and reduce the quality and immersiveness of the setup. Hence, it could be better to let the instructor or another human to reply on radio on behalf of the autonomous vessel, but it could be a challenge since it would be impossible for a human to accurately predict the algorithm's intentions and relay these truthfully on radio upon request.

5. Technical requirements

In order to use a maritime simulator as a test bed for autonomous vessels, it is crucial that the simulator provides an interface that enables third party computers to control various settings and controls in the simulation environment. Below, we have listed some interface properties in order of importance:

- controlling target vessel's speed and heading,
- reading position of fixed and moving obstacles (incl. other vessels),
- controlling (emulating) bridge handles and controls,
- reading bridge view (screen), radar, ECDIS, AIS and other instruments,
- controlling other simulation parameters, such as environmental settings, etc.

Since the designers of the maritime simulators probably did not have this use of the simulators in mind when the simulators were made, there is likely no dedicated interface readily available for this use. However, the features listed above correspond quite well with the needs for the operator/instructor's needs and are probably available through some API or similar used by the software for the operator station. However, existing APIs/interfaces are generally not public and most likely internal to the developer and subject to change. Hence, a crucial element in setting up such a proposed testing environment is access to such an interface, and this most likely depends on close cooperation with and goodwill from the simulator vendor.

6. Test setup and research questions

Below, we have listed two main test setups (see Fig. 6) for maritime training simulators and described what kind of interesting research questions these tests could contribute towards answering.

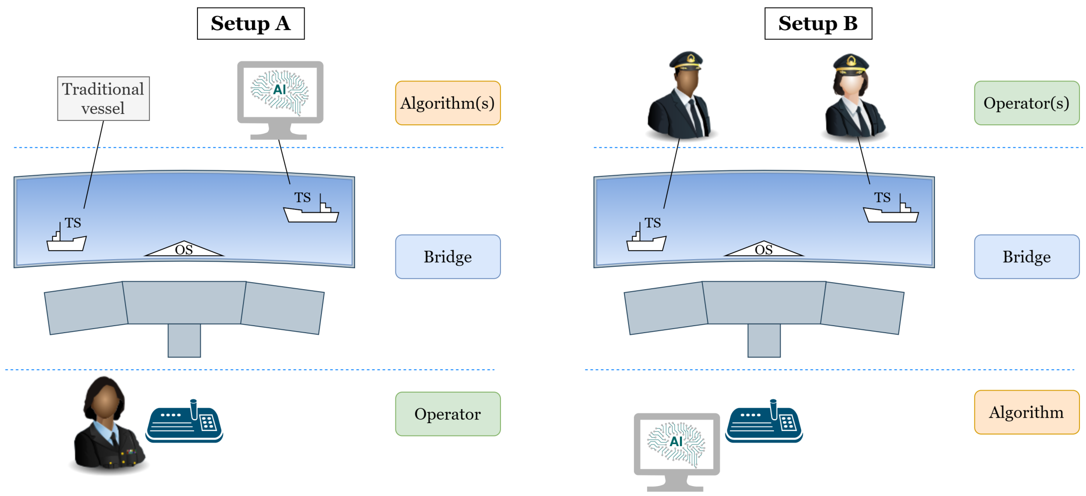

Figure 6. Visual representation of the test setups A and B.

- *Test setup A* — where the algorithm(s) control the target vessel(s) while the bridge(s) is/are controlled by human cadet(s).
- *Test setup B* — where the algorithm(s) control the bridge(s).

6.1. Test Setup A

In the first setup (Fig. 6, left), the algorithm(s) control the target vessel(s) and the bridge(s) is/are controlled by human cadet(s) / operator(s). In a mixed environment where traditional ships and autonomous ships co-exist, it is crucial to understand how they may interact. In this test, it is possible to study both how the autonomous vessel(s) responds to the human actions, and how humans react to the actions performed by the algorithm(s). In addition to evaluating the above-mentioned metrics, the cadets may be interviewed after the exercise in order to capture their experience with the autonomous vessel. This setup is suitable to test algorithms similarly to Turing's imitation game.

Another interesting research question is whether it is beneficial for navigators on other vessels to know that the vessel is autonomous? And further, would it be beneficial to know how the algorithm is "thinking"? With explainable AI, it could be possible for an autonomous vessel to broadcast its reasoning and intentions to other seafarers. Could this information be useful, or would it be a distraction? Could it contribute to information overload on humans? On the other side, would other autonomous vessels be able to perform better if they knew the intentions of other autonomous vessels, or could deadlock situations (more easily) occur?

6.2. Test Setup B

In the second setup (Fig. 6, right), the algorithm(s) control the bridge(s). One series of tests that can be performed is to run basic tests of an algorithm with respect to collision avoidance (avoiding fixed and moving obstacles) and even several vessels with identical or different algorithms. These tests can be done with more "Lo-Fi" simulators also, however, these certified simulators may give more reliable results.

A more interesting test is to give the algorithms the same test scenarios that the instructors give the cadets. The instructors have a set of test scenarios they run as part of their training of cadets. An experienced instructor has an excellent reference base on how well humans (cadets) solve the test scenarios. Hence, the instructors should have no problem evaluating how well the algorithm perform compared to humans using the same metrics as they normally do with the cadets. With some minor modifications, it could also be possible to do this test blindly, but the value of that should be weighted against the increased complexity.

Adding another layer of complexity to this test setup, the instructors could compare how well an algorithm performs compared to humans in a situation that could give humans a cognitive overload. One hypothesis could be that humans will outperform the computer in uncomplicated scenarios, however, the computer would be better at handling very complex situations. In addition to answering the aforementioned research questions, the results could also give some insight on the use of algorithms not only to navigate autonomous vessels, but also as decision support tools or as "smart autopilots" on manned vessels.

6.3. Other setups

The suggested setups and research questions above are of course just examples. It is possible to construct numerous different setups by introducing other elements, combining scenarios, etc. Advanced simulators like the maritime training simulators are a great tool that make it possible to do research with humans in the loop and with varying degree of complexity to validate autonomous ship algorithms.

7. Discussion

In order to do studies with humans in the loop, it is important to make a test scenario that is as close to reality as possible in a safe, controllable and observable environment. Although no simulators are exact copies of reality, maritime training simulators are deemed to be good training environments by international bodies such as International Maritime Organization (IMO) [45, 20]. Likewise, the cadets that participate in training are not experienced sailors, and therefore it is hard to generalise all findings on cadets to experienced sailors. However, this may be rectified by at least running a control group, where experienced sailors are subject to the same tests as the cadets and the algorithms.

In the future, autonomous vessels may be required to identify themselves as "robots" by transmitting some kind of identification to other seafarers. In this paper, we suggest that testing should be done also without mentioned identification. This is motivated by an assumption that in a mixed environment with regular and autonomous vessels, the traffic will flow safer and more efficiently if the autonomous vessels behave like regular vessels. We believe the proposed experiments will help shed light on this research question as well.

Additionally, in order to encourage and speed up the development of simulation-based testing for ship navigation algorithms, sharing open access data, software and results would be an important driver. However, it is also vital to consider data sharing restrictions from partners when they prevail.

All in all, we argue that test setups such as those described above will be able to provide valuable answers to research questions that are hard, if not impossible, to answer otherwise.

8. Conclusions

In this article, we have advocated for using maritime simulators to investigate how well algorithms for autonomous vessels perform in general and especially with humans in the loop. We have identified several interesting research questions that may be hard to answer without the use of the mentioned simulators, and we show that these simulators open up numerous possible research setups that may be explored. We have also identified the need for an interface that allows a third party computer to interact with the simulator. The need for test metrics has also been outlined.

Finally, we strongly argue that it is of great importance that the human-machine interaction between autonomous and traditional vessels is investigated in depth in order to prepare for a future where autonomous and traditional vessels will operate side by side.

Acknowledgments

We wish to thank Terje Ole Slinning and the Department of Ocean Operations and Civil Engineering at NTNU in Ålesund for providing photos and giving us a tour of their maritime training simulator area at the Norwegian Maritime Competence Center.

References

[1] International Maritime Organization 1972 COLREGs - International Regulations for Preventing Collisions at Sea
[2] Vagale A, Oucheikh R, Bye R T, Osen O L and Fossen T I 2021 Path planning and collision avoidance for autonomous surface vehicles I: a review *Journal of Marine Science and Technology (Japan)* **26** 1307–1323 ISSN 09484280
[3] Pedersen T A, Glomsrud J A and Haugen O I 2019 Towards simulation-based verification of autonomous navigation systems *International Seminar on Safety and Security of Autonomous Vessels* (Helsinki) pp 1–13
[4] Wood M, Robbel P, Maass M, Tebbens R D, Meijs M, Harb M, and Schlicht P 2019 Safety first for automated driving
[5] Spanfelner B, Richter D, Ebel S, Wilhelm U, Branz W and Patz C 2012 Challenges in applying the ISO 26262 for driver assistance systems *Tagung Fahrerassistenz, München* **15** 2012
[6] Koopman P and Wagner M 2016 Challenges in autonomous vehicle testing and validation *SAE International Journal of Transportation Safety* **4** 15–24
[7] Brandsæter A and Knutsen K E 2018 Towards a framework for assurance of autonomous navigation systems in the maritime industry *Safety and Reliability–Safe Societies in a Changing World : Proceedings of ESREL 2018* (CRC Press) pp 449–457
[8] Kongsberg Digital 2021 K-SIM® Navigation URL https://www.kongsberg.com/digital/products/maritime-simulation/k-sim-navigation/
[9] Wärtsilä 2022 Wärtsilä Navigation Simulator NTPRO 5000 URL https://www.wartsila.com/voyage/simulation-and-training/ntpro-5000-simulator
[10] VSTEP NAUTIS Navigation Simulator URL https://www.vstepsimulation.com/nautis-simulator/nautis-maritime-simulator/navigation/
[11] Endsley M R 2019 Level of automation effects on performance, situation awareness and workload in a dynamic control task *Hearing on Boeing 737-Max8 Crashes — December 11, 2019*
[12] Heierstad T 2020 Simulation technology – a vital resource for testing autonomous ship operations - Kongsberg Digital URL https://www.kongsberg.com/digital/resources/stories/2020/6/simulation-technology--a-vital-resource-for-testing-autonomous-ship-operations/
[13] Wärtsilä 2020 Making ship systems smarter with simulation URL https://www.wartsila.com/insights/article/making-ship-systems-smarter-with-simulation
[14] Open Simulation Platform 2022 Open Simulation Platform. Towards a maritime ecosystem for efficient co-simulation URL https://opensimulationplatform.com/

[15] Vagale A, Bye R T and Osen O L 2020 Evaluation of Path Planning Algorithms of Autonomous Surface Vehicles Based on Safety and Collision Risk Assessment *OCEANS 2020 MTS/IEEE Global: Singapore - U.S. Gulf Coast* pp 1–8

[16] IMO 2011 International Convention on Standard of training, certification and watchkeeping for seafarers, including 2010 Manila Amendments

[17] Brandsæter A and Osen O L 2021 Assessing Autonomous Ship Navigation using Bridge Simulators Enhanced by Cycle-Consistent Adversarial Networks *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* 1–9

[18] Dahlstrom N, Dekker S, Van Winsen R and Nyce J 2009 Fidelity and validity of simulator training *Theoretical Issues in Ergonomics Science* **10** 305–314 ISSN 1464536X

[19] Barnett M, Gatfield D and Pekcan C 2003 A Research Agenda in Maritime Crew Resource Management *Proceedings of the International Conference on Team Resource Mangement in the 21st Century* pp 1–22

[20] DNV 2017 Maritime Simulator Systems *Standard DNVGL-ST-0033* URL https://rules.dnvgl.com/docs/pdf/DNVGL/ST/2017-03/DNVGL-ST-0033.pdf

[21] FORCE Technology 2014 Ship bridge simulators URL https://forcetechnology.com/en/all-industry-facilities/ship-bridge-simulators

[22] Poseidon Simulation AS Borealis Navigation Simulator URL https://poseidon.no/products/borealis/

[23] BMT BMT REMBRANDT: high integrity ship simulations URL https://www.bmt.org/our-innovations/bmt-rembrandt/

[24] ARI Simulation 2021 Ship Maneuvering Simulators URL https://arisimulation.com/products/marine-simulators/bridge

[25] Image Soft 2022 IS Full Mission Bridge Simulator URL https://imagesoft.fi/product/is-full-mission-bridge-simulator-for-certified-maritime-training/

[26] Rheinmetall AG 2022 Bridge and navigation training URL https://www.rheinmetall-defence.com/en/rheinmetall_defence/systems_and_products/simulation_and_training/civil_training_solutions/merchant_marine_training/bridge_and_navigation_/index.php

[27] DNV 2018 Class Guideline: Autonomous and remotely operated ships

[28] International Standardization Organization 2018 ISO 26262-1:2018(en) Road vehicles — Functional safety

[29] Salay R, Queiroz R and Czarnecki K 2018 An analysis of ISO 26262: Using machine learning safely in automotive software *arXiv preprint arXiv:1709.02435*

[30] Caruana R, Kangarloo H, Dionisio J D, Sinha U and Johnson D 1999 Case-based explanation of non-case-based learning methods. *Proceedings of the AMIA Symposium* (American Medical Informatics Association) p 212

[31] Doshi-Velez F and Kim B 2017 Towards A Rigorous Science of Interpretable Machine Learning *arXiv: Machine Learning* 1–13 URL http://arxiv.org/abs/1702.08608

[32] Lundberg S M and Lee S I 2017 A unified approach to interpreting model predictions *Advances in Neural Information Processing Systems* pp 4765–4774

[33] Brandsæter A, Smefjell G, van de Merwe K and Kamsvåg V 2020 Assuring Safe Implementation of Decision Support Functionality based on Data-driven Methods for Ship Navigation *I: e-proceedings of the 30th European Safety and Reliability Conference and 15th Probabilistic Safety Assessment and Management Conference (ESREL2020 PSAM15)*

[34] Liu M Y, Breuel T and Kautz J 2017 Unsupervised image-to-image translation networks *Advances in neural information processing systems* pp 700–708

[35] Zhu J Y, Park T, Isola P and Efros A A 2017 Unpaired image-to-image translation using cycle-consistent adversarial networks *Proceedings of the IEEE international conference on computer vision* pp 2223–2232

[36] Pei K, Cao Y, Yang J and Jana S 2017 DeepXplore: Automated Whitebox Testing of Deep Learning Systems *arXiv preprint arXiv:1705.06640*

[37] Pei K, Cao Y, Yang J and Jana S 2017 Towards Practical Verification of Machine Learning: The Case of Computer Vision Systems *ArXiv* **1712.01785** 1–16

[38] Endsley M R 1988 Design and Evaluation for Situation Awareness Enhancement: *Proceedings of the Human Factors Society Annual Meeting* vol 32 (SAGE PublicationsSage CA: Los Angeles, CA) pp 97–101

[39] Endsley M R 1988 Situation Awareness Global Assessment Technique (SAGAT) *IEEE Proceedings of the National Aerospace and Electronics Conference* (IEEE) pp 789–795

[40] Endsley M R 2000 Direct Measurement of Situation Awareness: Validity and use of SAGAT *Situation Awareness Analysis and Measurement*

[41] Øvergård K I, Nazir S and Solberg A S 2017 Towards automated performance assessment for maritime navigation *TransNav, International Journal on Marine Navigation and Safety of Sea Transportation* **11** 229–234

[42] Vagale A, Bye R T, Oucheikh R, Osen O L and Fossen T I 2021 Path planning and collision avoidance

for autonomous surface vehicles II: a comparative study of algorithms *Journal of Marine Science and Technology (Japan)* **26** 1307–1323 ISSN 09484280
[43] Naeem W, Irwin G W and Yang A 2012 COLREGs-based collision avoidance strategies for unmanned surface vehicles *Mechatronics* **22** 669–678 ISSN 09574158
[44] Wang N, Gao Y, Zheng Z, Zhao H and Yin J 2018 A Hybrid Path-Planning Scheme for an Unmanned Surface Vehicle *8th International Conference on Information Science and Technology* pp 231–236
[45] Zghyer R and Ostnes R 2019 Opportunities and Challenges in Using Ship-Bridge Simulators in Maritime Research *Proceedings of Ergoship 2019* (Haugesund: Western Norway University of Applied Sciences (HVL)) pp 119–131