Human Factors in Robots, Drones and Unmanned Systems, Vol. 93, 2023, 76–84
https://doi.org/10.54941/ahfe1003750

Decision Transparency for Enhanced Human-Machine Collaboration for Autonomous Ships

Andreas Nygard Madsen[1], Andreas Brandsæter[1,2], and Magne V. Aarset[1,3]

[1]Norwegian University of Science and Technology, Alesund, Norway
[2]Volda University College, Volda, Norway
[3]TERP AS, Haugesund, Norway

ABSTRACT

Maritime Autonomous Surface Ships (MASS) are quickly emerging as a game-changing technology in various parts of the world. They can be used for a wide range of applications, including cargo transportation, oceanographic research and military operations. One of the main challenges associated with MASS is the need to build trust and confidence in the systems among end-users. While the use of AI and algorithms can lead to more efficient and effective decision-making, humans are often reticent to rely on systems that they do not fully understand. The lack of transparency and interpretability makes it very difficult for the human operator to know when an intervention is appropriate. This is why it is crucial that the decision-making process of MASS is transparent and easily interpretable for human operators and supervisors. In the emerging field of eXplainable AI (XAI), various techniques are developed and designed to help explain the predictions and decisions made by the AI system. How useful these techniques are in a real-world MASS operation is, however, currently an open question. This calls for research with a holistic approach that takes into account not only the technical aspects of MASS, but also the human factors that are involved in their operation. To address this challenge, this study employs a simulator-based approach were navigators test a mock-up system in a full mission navigation simulator. Enhanced decision support was presented on an Electronic Chart Display & Information System (ECDIS) together with information of the approaching ships as AIS (Automatic Identification System) symbols. The decision support provided by the system was a suggested sailing route with waypoints to either make a manoeuvre to avoid collision, or to maintain course and speed according to the Convention of the International Regulations for Preventing Collisions at Sea (COLREG). After completing the scenarios, the navigators were asked about the system's trustworthiness and interpretability. Further, we explored the needs for transparency and explainability. In addition, the navigators gave suggestions on how to improve the decision support based on the mentioned traits. The findings from the assessment can be used to develop a strategic plan for AI decision transparency. Such a plan would help building trust in MASS systems and improve human-machine collaboration in the maritime industry.

Keywords: Decision transparency, Mass, XAI, Human-machine collaboration, Decision support

INTRODUCTION

Imagine you are driving a car. It is an advanced vehicle with a high degree of automation. Suddenly it tells you to move into the opposite lane, and you don't understand why. Would you do it? Most people would not, since we humans are reticent to apply and trust in decision support that we do not fully understand (Aarset and Johannessen, 2022). With Maritime Autonomous Surface Ships (MASS) on the horizon, both industry and academia are focusing on autonomous collision avoidance systems, sensors and decision support systems (DSS). As the development and implementation of autonomous systems continue to advance, there is a growing concern about the transparency and interpretability of such systems. Some models are simple by design and can easily be interpreted by human users. Others are extremely complex and complicated. The algorithms used to make the models can be simple to understand and implement, but after training, the final models can become very complex and may be impossible to understand and interpret (Brandsaeter and Glad, 2022).

To overcome this problem, researchers are developing various methods to help explain the "reasoning" of the system. In computer science this problem area is referred to as eXplainable Artifical Intelligence (XAI). Explainability refers to any action taken by an AI with the intent of clarifying its internal functions (Barredo Arrieta et al., 2020). In other words, a model can be explained using methods and tools from XAI. Doshi-Velez and Kim, (2017) define interpretability as "the ability to explain or to present in understandable terms to a human.". For a navigator onboard a ship or in a Remote Operation Centre (ROC) working together with an AI, it is important that the AI is transparent about how it "thinks", and that the reasoning behind any decision is transparent. It is also important that alternative decisions are easily available for the navigator. With regards to autonomous navigation and collision avoidance, this issue can be termed AI Decision Transparency. This approach introduces a focus on how information from such a system should be presented to a navigator, regardless of the system is doing decision-making or providing decision support.

The International Maritime Organization (IMO) has recognized the potential of MASS and described different levels of MASS based on degree of autonomy. These levels are characterized by the degree of automation employed, with some systems requiring minimal human intervention, while others rely heavily on human input. In either case, the importance of transparency and interpretability cannot be overstated.

EXPERIMENT DESIGN AND METHOD

This paper aims to explore the need for decision transparency in decision support systems for collision avoidance in ship navigation. This is based on inspiration from novel industry systems, where decision support is presented on an ECDIS or similar. This is a reasonable starting point, since the end-users (navigators) are familiar with ECDIS, and it contains valuable information such as the nautical chart and AIS targets. Furthermore, the display is not affected by clutter and noise, which could be the case in radar systems.

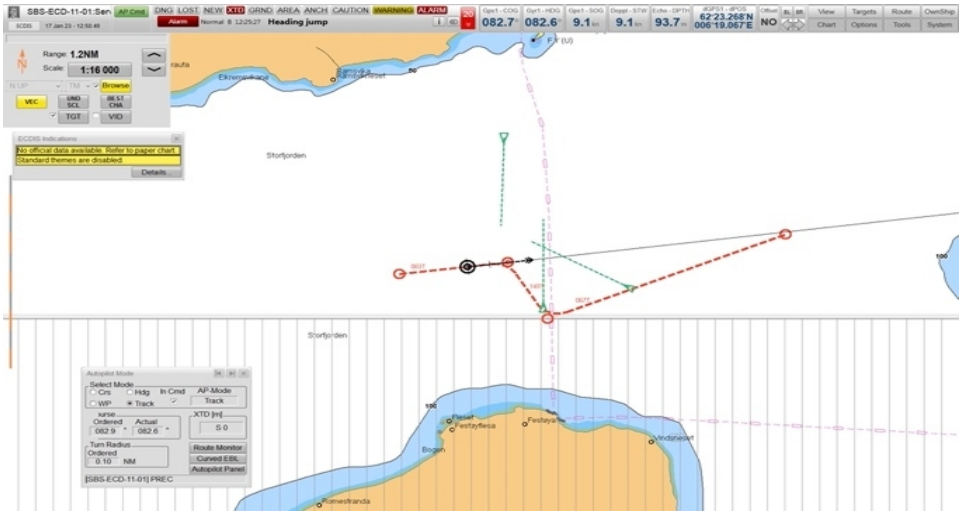

Figure 1: Decision support presented on secondary ECDIS. The ownship is represented by the "double circle", and the dotted red line is the suggested manoeuvre to avoid collision.

The participants in this simulator-based experiment were seven senior year deck officer students, where all students have a minimum of two years of experience at sea. They were presented with what we consider minimal explainability; a proposed sailing route to avoid collision or close encounters. This way, it is possible to see if the situation in itself is interpretable and enables the navigators to understand the system's suggestions.

Simulator

The experiment utilize a full mission bridge simulator. This aids in comparing and assessing how functionality for autonomous navigation and decision support affects navigation performance and safety (Brandsæter and Osen, 2021; Mislevy, 2013). Simulators allow to perform controlled, repeated experiments with identical initial conditions, which facilitates for assessment and comparing of how different candidates respond to the decisions of the system.

The simulator has a visual view of 120 degrees, autopilot, throttle, rudder control, radar and two ECDIS.

Scenario

The participants were briefed that they were situated in a remote operation center (ROC), and in charge of eight vessels. Further, they were informed that they did not have the capacity of knowing the details of every ship's situation, but they knew where the ships were. In this way, the participants were simulated as out of the loop. The ships were operated by an AI (mock-up), and if it encountered a situation where the system required a navigator to verify its intention, the system would raise a red flag and the navigator had to attend to the bridge. When a participant attended the bridge, there was a

Figure 2: The layout of the ship bridge simulator (Kongsberg Digital, 2020).

minimum of 60 seconds until a maneuver had to be performed to avoid collision. The participants' task was to assess the situation, using the equipment on the bridge, and resolve the situation, either by following the information presented by the DSS, or by a maneuver of their own choice. Each participant completed eight scenarios.

Decision Support/Mock-Up AI

When designing the scenarios, a proposed maneuver to avoid collision was decided by an experienced navigator and transferred to one of the two ECDIS in the bridge simulator. The autopilot was set to track control, which means the ship would follow the pre-planned maneuver unless the navigators intervened. In four out of eight scenarios the decision support provided was sub-optimal, and the participants should in these cases preferably deviate from the suggestions from the DSS. In these cases, the sub-optimal decision support would not make the vessels collide, but there were indeed more optimal solutions to the situations.

Traffic Situations

The traffic situations in the scenarios were based on real and documented situations in a fjord in western Norway. In all the scenarios, the ship is in the same geographical area, crossing a coastal ferry route. The data and data collection process is presented by Rutledal et al. (2020), and also in Madsen et al. (2022).

Method

After completing eight scenarios, the participants filled in a questionnaire (Appendix I) regarding interpretability and trustworthiness for this way of displaying information. The questionnaire used linear scales (e.g., Very bad Very good), and the participants made a mark on the scale. This has later

Figure 3: Geographical area for the scenarios.

been transformed to the scale 0-100%. Afterwards, the navigators were interviewed based on their questionnaire answers. Further, the participants were given the opportunity to draw and explain how they would prefer a system to display information to enhance the decision transparency of such a system.

RESULTS AND DISCUSSION

The navigators were asked how trustworthy, accurate and reliable they found the suggestions and information from the system to be. Note that the participants were unaware of the system being a mock-up. Table 1 shows the different candidates score on system accuracy, system reliability and system trustworthiness. For example, an accuracy of 50% means that the candidate made a mark in the middle of the scale between very bad and very good. Each result is based on a single question.

Table 1 present results of the navigators' assessment of the system. Most of the participants found the system to be above 50 % accurate, reliable and trustworthy on the scale 0–100 %. Only one of the participants found the system's reliability to be less than 20 %. As mentioned above, in four of the scenarios the suggestions from the system was sub-optimal. When the participants were asked if there were any instances where they felt the system's suggestions were misleading, two of the participants answered "No", two answered "Yes", two answered "Yes, the system should start its manoeuvre sooner", and two answered "Yes" and referred to the scenario" which is illustrated in Figure 4.

When the participants were asked about how comfortable they would feel to rely on the system for critical decisions, there is a much lower score. This

Table 1. Results from the simulations wrt accuracy, reliability, and trustworthiness.

Participant no.	1	2	3	4	5	6	7	Mean
Accurate	74 %	86 %	51 %	88 %	68 %	20 %	36 %	60 %
Reliable	71 %	82 %	19 %	98 %	62 %	47 %	49 %	61 %
Trustworthy	66 %	71 %	32 %	70 %	53 %	51 %	64 %	58 %

Table 2. Results from the simulations wrt comfortability.

Participant no.	1	2	3	4	5	6	7	Mean
Critical decisions	68 %	75 %	19 %	45 %	31 %	51 %	16 %	44 %

Table 3. Results from the simulations wrt understandability.

Participant no.	1	2	3	4	5	6	7	Mean
Understanding of reasoning for system suggestions	74 %	78 %	46 %	72 %	64 %	21 %	59 %	59 %

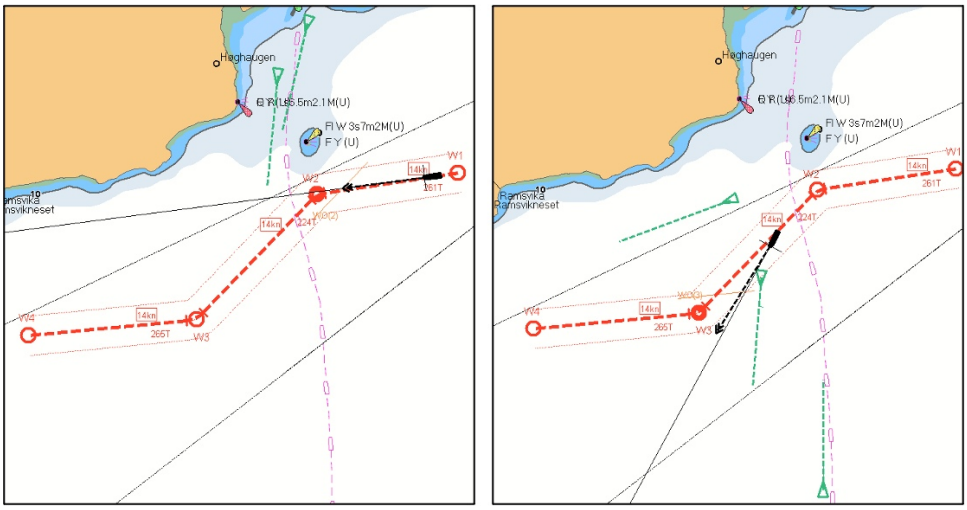

Figure 4: Decision support before and after manoeuvre. The picture on the left hand side shows the situation 10 seconds into the experiment. The picture on the right hand side shows the situation after 4 minutes.

is not unexpected as the DDS performs suboptimal in four out of eight scenarios. What is, perhaps, more surprising is that the numbers are this high. The results are shown in Figure 2.

Table 3 shows how the participants answered question "How easy was it for you to understand how the system arrived at its suggestion?"

We see that the navigators think that they are able to understand the reasoning behind the system to some extent (mean 59 %). In the interviews this is explained that the solution to the situation is self-explanatory by observing the traffic and consulting the traffic regulations (COLREG). In the interviews, five of the participants say that the reasoning and interpretability of the suggestions is much harder to comprehend when there are more than one ship taken into consideration by the system. Almost all of the participants refer to the situation in Figure 4.

Figure 4 shows a situation where two ships are approaching from the starboard side. The suggestion from the system is sub-optimal, but the participants are not able to see the intent of the system. In the interviews, the participants explains that they it appears that the system suggests turning to port and pass ahead of both ships. If the ownship follow the proposed route, it will actually pass astern of the first target, and then turn to port to increase the CPA of the second target. Three of the participants referred to this scenario when answering to the question "Were there any instances where the systems suggestions was unclear and ambiguous".

Results from the simulations with respect to understanding and situational awareness (SA) is presented in Table 4. We see that the participants rank the importance of understanding the reasoning behind the systems suggestions quite high (mean 87 %), but there is a low understanding of how the system acquire its SA (mean 40 %). All navigators believe that the system should provide information about how it acquires its SA.

The participants were encouraged to both draw and explain how they think that such a system might be improved, to enhance the transparency of the systems decisions and in turn their own SA. All of the participants called for a "trial function", where they could fast forward the situation to see the outcome of the manoeuvre. An on-demand animation was also proposed, with one window illustrating how the system think the situations will be resolved with the suggested decision, and a second window illustrating how the system believe the situation would play out if no action is taken. Three of the participants suggested that the system should indicate which target ships in a situation was taken into consideration when suggesting a manoeuvre. Furthermore, four participants called for information about which COLREG rules the system utilized, and information regarding why the system believe that this is the correct manoeuvre (e.g. type of target ship and aspect). Two of the participants answered that the system should provide multiple actions, in a prioritized order.

Nevertheless, the navigators overall experience of their interaction with the system is quite good as shown in Table 5.

Table 4. Results from the simulations wrt understanding and situational awareness (SA).

Participant no.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Mean**
Importance of understanding the reasoning behind the system	91 %	98 %	65 %	100 %	85 %	97 %	73 %	87 %
Understanding of system's ability to build SA	56 %	19 %	32 %	11 %	51 %	69 %	40 %	40 %
Should the system provide info about how it builds SA?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	

Table 5. Results from the simulations wrt overall experience.

Participant no.	1	2	3	4	5	6	7	Mean
Overall experience of interaction with system	82 %	88 %	72 %	73 %	73 %	59 %	70 %	74 %

In the interviews, all of the participants conclude that this kind of system would be helpful, even though they don't necessarily agree with the systems suggestions. To quote one of the navigators: "*This sort of system would be great to have onboard. It might not provide the best solution but it gives you an indication of the danger and is really helpful as a starting point*".

CONCLUSION

This paper have discussed how to provide input to and uncover the needs for decision transparency in decision support systems for collision avoidance, based on some preliminary simulations. The results indicate that a system must be able to communicate its suggestions to the user in such a way that the decisions are transparent and that alternative decisions are easy to execute. Furthermore, the results strongly indicates that the system must provide information about how it builds its SA. This indicates that further research in developing strategies for AI decision transparency is needed. Based on the experience from this study, there is a need to focus on how a decision support system can be part of a resilient integrated system to ensure and maintain a reliable situational awareness.

FUNDING

This work has been funded by the Research Council of Norway: Center for Research-based Innovation *AutoShip* (project number 309230) and *SAFE Maritime Autonomous Technology* (project number 327903).

ACKNOWLEDGEMENTS

The authors would like to thank all participants, and the Vocational College Møre & Romsdal for the use of their simulators.

REFERENCES

Aarset, M. V., Johannessen, L. K., 2022. On Distributed Cognition While Designing an AI System for Adapted Learning. Front Artif Intell 5. https://doi.org/10.3389/FRAI.2022.910630

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., Herrera, F., 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion 58, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Brandsæter, A., Osen, O. L., 2021. Assessing autonomous ship navigation using bridge simulators enhanced by cycle-consistent adversarial networks. Proc Inst

Mech Eng O J Risk Reliab. https://doi.org/10.1177/1748006X211021040/ASSET/IMAGES/LARGE/10.1177_1748006X211021040-FIG6.~JPEG.

Brandsaeter, Andreas, Glad, Ingrid K, 2022. Shapley values for cluster importance. Data Min Knowl Discov 1–32. https://doi.org/10.1007/s10618-022-00896-3

Doshi-Velez, F., Kim, B., 2017. Towards A Rigorous Science of Interpretable Machine Learning.

Madsen, A. N., Aarset, M. V., Alsos, O. A., 2022. Safe and efficient maneuvering of a Maritime Autonomous Surface Ship (MASS) during encounters at sea: A novel approach. Maritime Transport Research 3, 100077. https://doi.org/10.1016/J. MARTRA.2022.100077

Mislevy, R. J., 2013. Evidence-Centered Design for Simulation-Based Assessment. Mil Med 178, 107. https://doi.org/10.7205/MILMED-D-13-00213

Rutledal, D., Relling, T., Resnes, T., 2020. It's not all about the COLREGs: A case-based risk study for autonomous coastal ferries. IOP Conf Ser Mater Sci Eng 929. https://doi.org/10.1088/1757-899X/929/1/012016