Proceedings of the 32nd European Safety and Reliability Conference (ESREL 2022)
Edited by Maria Chiara Leva, Edoardo Patelli, Luca Podofillini, and Simon Wilson

Published by Research Publishing, Singapore.
doi: 10.3850/978-981-18-5183-4_S05-01-186-cd

Approvable AI for Autonomous Ships: Challenges and Possible Solutions

Brian Murray
SINTEF Ocean, Norway. E-mail: brian.murray@sintef.no

Ørnulf Jan Rødseth
SINTEF Ocean, Norway

Håvard Nordahl
SINTEF Ocean, Norway

Lars Andreas Lien Wennersberg
SINTEF Ocean, Norway

Armin Pobitzer
SINTEF Ålesund, Norway

Henrik Foss
Kongsberg Seatex, Norway

Artificial Intelligence (AI) is being promoted as an important contributor to ensuring the safety of autonomous ships. However, utilizing AI technologies to enhance safety may be problematic. For instance, AI is only capable of performing well in situations that it has been trained on, or otherwise programmed to handle. Quantifying the true performance of such technologies is, therefore, difficult. This raises the question if these technologies can be applied on larger ships that need approval and safety certification. The issue gains further complexity when introduced as an element in remote control centres. This paper presents an overview of the most relevant applications of AI for autonomous ships, as well as their limitations in the context of approval. It is found that approval processes may be eased by restricting the operational envelope of such systems, as well as leveraging recent developments in explainable and trustworthy AI. If leveraged properly, AI models can be rendered self-aware of their limitations, and applied only in low risk situations, reducing the workload of human operators. In high risk situations, e.g. high AI model uncertainty or complex navigational situations, a timely and effective handover to a human operator should take place. In this manner, AI-based systems need not be capable of handling all possible situations, but rather be capable of identifying their limitations and alerting human operators to situations that they are incapable of handling with an acceptable level of risk.

Keywords: Autonomous ships, Artificial Intelligence, Machine Learning, Approval and certification

1. Introduction

The field of autonomous ships is developing rapidly. Norway is arguably at the forefront of much of the development in autonomous ships, with test areas outside Trondheim, Horten, Haugesund, and Ålesund. Furthermore, Massterly, the world's first operator of autonomous vessels, will be operating vessels for Yara and ASKO, initially with crew onboard, but with increased levels of automation introduced over the coming years, until 2024 when these ships are planned to be partly autonomous and operrated with no crew onboard. Thus, there is an urgent need to establish practically useful approval procedures.

Regulatory bodies are currently attempting to establish guidelines for approval of autonomous ships. Most approaches are largely based on the IMO circular MSC.1/Circ.1455 (IMO, 2013) e.g. Norwegian Maritime Authority (2020) DNV (2021a) and BV (2019). These procedures are generally based on a systematic risk assessments of the planned operations and the equipment in use, to ensure a safety level equivalent to that on conventional ships. However, autonomous ships will likely utilize Artificial Intelligence (AI)-based systems, an element that the guidelines do not address in detail. DNV (2021b) addresses assurance of data-driven models, but lacks the bridge to approval of AI-based systems in the context of autonomous shipping. Hence, such systems will require verification of system performance as the guidelines currently stand, resulting in a number of challenges.

AI has a highly variable status in autonomous systems research. Some authors claim that a system cannot be fully autonomous unless it includes technology that allows the system to learn while acting. For example,Williams and Scharre (2015) refer to level 4 autonomy defined by a NATO's Industrial Advisory Group (NIAG) group as *"autonomous learning system with the ability to modify rule-defining behaviors"*. This would imply using AI to, e.g. improve the system's observation and decision-making capabilities. On the other hand, this use of AI would change the system's behavior over time, which means that the system after some time of operation would not be the same system as when it started. This will be very problematic for systems that require approval or certification to operate, e.g. ships. Approval processes in current guidelines aim to verify that the system capabilities, when the system is put into operation, are sufficiently well implemented to handle all relevant situations that the system will encounter and that there are no aspects of the system's construction that can cause significant safety risks. If the system changes behavior over time, it will be a challenge to prove that this does not cause new and possibly unknown safety risks (Koopman and Wagner, 2017). Currently, our opinion is that including learning as part of safety certified autonomous vehicle's capabilities is not realistic.

This does not preclude using AI in system components. In the context of autonomous ships there are two particular areas where the use of AI is interesting: 1) In detection and classification, most commonly in various forms of environment perception such as scenery and objects, but also for condition monitoring of equipment, events and states; 2) In prediction of future situations and in planning suitable actions, such as collision avoidance. The use of AI in both these areas should be feasible given that the functionality of the AI system can be verified for the intended use. This paper aims to provide an overview of AI for autonomous ships, as well as the challenges in the context of approval processes. Furthermore, possible ways to ameliorate approval process are presented, as a step towards aiding future regulatory guidelines.

2. What is AI for autonomous ships?

2.1. *What is AI?*

AI makes use of a number of different technologies and this paper will not try to give full classification of these. Since the 1990s, statistical methods and in particular various forms of artificial neural networks (ANN) and machine learning (ML) have re-emerged as the most important area of AI applications. Machine learning differs from "Good Old-Fashioned AI" (GOFAI) (Haugeland, 1989), e.g. symbolic and fuzzy logic, in that its response to stimuli is not any more explainable by an explicitly specified computer code or database.

Most ANNs that are used today are deep neural networks, i.e. they contain several "hidden" layers in addition to the input and output layers. ML techniques are typically categorized into algorithms for supervised learning, unsupervised learning and reinforcement learning (Al Ridhawi et al., 2020). Many of these methods are based on the training of an ANN with data, whereby the network establishes an internal weighting of different characteristic factors of the input data sets, e.g., to detect certain patterns or to give suggestions for suitable actions. ANNs provides state-of-the-art performance within computer vision, allowing for the detection and classification of objects e.g. ships. Furthermore, it can be used for regression tasks by mapping non-linear relationships that can be used for predictions, e.g. the future path of a vessel.

2.2. *Where in the ship can AI be used?*

An autonomous ship system consists of many components located both on the ship and remotely from the ship, such as in a Remote Control Center (RCC). AI can be used in many of these components, but in this paper, we will discuss two specific applications. 1) The Situation Awareness System (SAS), i.e. the sub-system that acquires data about the ship and its surroundings and builds a computer model of objects that can have an impact on the operation of the autonomous ship (i.e. detection and classification); and 2) The Autonomous Navigation System (ANS), i.e. the system that controls the ship's movements to ensure a safe passage, including track and speed control as well as collision avoidance (i.e. evaluating future situations and planning suitable actions).

Both the SAS and ANS functions have been extensively researched, see e.g. Thombre et al. (2020) and Zhang et al. (2021). However, there is a fundamental difference between these two functions. A successful SAS is in principle only dependent on enough sensor data and sufficiently good algorithms to provide an arbitrary high quality environmental model. ANS, on the other hand, may ultimately have to make decisions based on what a human officer on the encountered ship will or will not do.

The Convention on the International Regulations for Preventing Collisions at Sea from IMO (1972), or COLREGs for short, is the internationally agreed on framework for how ships shall behave to avoid collisions. This includes the use of light and sound signals as well as guidelines for manoeuvres during encounters between ships in various situations and visibility conditions. However, these guidelines are generally only valid for two ships at a time and also refers to the use of ”good seamanship” in conjunction with collision avoidance. The actual actions taken in a given situation will therefore to a significant degree depend on the preferences of the bridge personnel. Thus, it is highly unlikely that the automation system, with or without AI, can handle all situations the ship may encounter, in particular situations where more than two ships meet or when another ship behaves erratically. Without some changes in the COLREGs, this is in principle not possible to do with 100% confidence (Rødseth et al., 2021).

The most common current assumption is that the autonomous ship will need a human operator to intervene in such situations. The automation system will, however, be able to handle the ship most of the time and one cannot rely on the operator to stay sufficiently alert during automatic operation to rapidly understand when it is necessary to take over control. One will need an alert from the automation system to the operator. This requires that the automation system and its AI can determine when its limits are approaching and alert the operator in time for safe handover. This can probably best be implemented by a rule-based system. AI is likely most suitable for use in SAS, where the most relevant applications of AI are object detection and classification, as the domain of these functions is fairly constrained.

3. Requirements to AI in SAS

The basic and most important requirement for the SAS in terms of collision avoidance, is that it can detect all objects with potential of causing a collision. Collision avoidance also depends on motion prediction for both the own ship and the ship (or object) which one may collide with, i.e. the target ship. That is, the future trajectory of own ship and the target ship must be predicted. Several methods for trajectory prediction are discussed in Huang et al. (2020). The SAS must, therefore, estimate the position, heading, and speed of the target ship as well as own ship. Such predictions will, however, be associated with significant uncertainty, as humans may not adhere to expected behavior in all cases.

The accuracy of the object detection, the state estimation (object speed, heading, etc.), and categorisation must always be estimated by the SAS such that the ANS can decide if the data is reliable enough for collision avoidance. An important part of this is continuous sensor integrity checking and verification. When the sensed environment cannot be determined with sufficient accuracy, the ANS must call upon the RCC for support. However, calling upon RCC should be minimized. This means that the SAS must be designed to function sufficiently well in all conditions the autonomous ship is expected to operate in.

4. Limitations of AI in SAS

The question on how to assess and approve AI and ML applications used on ships should not only target the specific algorithm or model, but instead encompass the complete model management of the data processing pipelines that are used (Schelter et al., 2018). Table 1 illustrates a set of generic steps associated with such a pipeline along with challenges that can be attributed to each step (Elshawi et al., 2019). The challenges that are listed are either generic (G) for ML and AI applications or application specific (A), i.e. SAS for object detection and classification, or autonomous ships.

Table 1. Data processing pipeline and the steps in developing ML solutions with reference to challenges listed in Table 2.

No.	Description	G	A
1	Data collection, cleansing, transformation, exploration	1,2,3	4,5
2	Choice of algorithm	6	7
3	(Re)Training, validation, test	8,9,10	
4	Evaluation of results	11,12	
5	Gap analysis	13,14	
6	Deploy/replace, versioning	15,16,17	17

The challenges are listed in Table 2 (Schelter et al., 2018; Feng et al., 2020; Moosbauer et al., 2019; Shin et al., 2020; Gupta et al., 2009; Ravindran et al., 2020). They are, furthermore, sorted into categories in Table 1, as items that developers (e.g. original equipment manufacturers) must be aware of to ensure desired model performance, that approval authorities (e.g. classification societies) must be aware of to be able to approve the use of a model, and finally that users (e.g. crew or RCC operators) need to be aware of to understand the constraints and limitations associated with the implementation and use of the model on an autonomous ship. The different stakeholders will have different involvement in the data processing pipeline, but the expectations to each stakeholder and each step are not defined as the regulatory frameworks for ships and inland waterways vessels do not pose explicit requirements to the use of AI and ML on autonomous ships, and furthermore

Table 2. Challenges associated with ML/AI applications that use ANN/DNN for SAS to detect and classify objects.

No.	Description
1	Skewed datasets/categories intended for classification models (many observations of few categories vs. few observations of many categories)
2	Data quality vs. data diversity
3	Labelling errors
4	Lack of public maritime benchmark datasets
5	No definition of common maritime ontology for object classification
6	Conventional classification methods assume independent and identical distribution (I.I.D. assumption)
7	Fusing strategy for multi-object detection
8	Accuracy improvements slows down in highly tuned ML models
9	Comparison of different models requires that the same data set is used for training, validation and test
10	Overfit on training and validation sets due to extensive experiments
11	Proper split of training, validation and test sets (I.I.D. assumption does not hold in real world applications)
12	Standardized metrics is needed for comparison of different models
13	Backtracking, i.e. to understand changes in model performance e.g. after re-training
14	Standardized metrics is needed to compare the performance of one or several ML model
15	Determine release candidates (old vs. new models, prediction times etc.)
16	Backwards compatibility of models
17	Determine and detect if and when retraining is needed

how to manage the model throughout the ship's life cycle.

This overview highlights two ML aspects of particular relevance for approvability in the context of autonomous vessels: firstly, that operational model performance is highly dependent on the available training data and training process, and, secondly, that performance is bound to change over its life cycle. The first point is important since it implies that the true operational performance and behavior of a model might be difficult to assess through a predefined set of test cases, or the like, while the second point raises the question of how a potential approval should be handled in the light of changes in model performance over time.

Hence, approval of systems involving ML-models diverges from general aspects of approvablility related to more regular equipment where behavior is guaranteed through the approval process. Within the domain of training data that the ML-model is trained on, performance can be quantified. Approval processes can implement standardized requirements to ML-models that can optimize performance within this domain, but many of the underlying challenges of utilizing ML-models will persist. By minimizing the effect of the challenges presented in this section, models will have a higher degree of operability. However, there is a large degree of uncertainty related to situations in which the model has limited experience.

Guaranteeing the performance of ML-models in all possible situations is, therefore, infeasible. Requirements could rather be set with respect to awareness of model limitations, as well as improved human-machine collaboration. Therefore, it is suggested in the next sections to leverage humans in conjunction with explainable AI to handle situations in which the relevant ML-models perform poorly.

5. Overcoming Limitations

5.1. *Human-AI Interaction*

The RCC is a central aspect of future autonomous ship systems. Hence, humans will provide a central component of the reliability of the system, as situations the ship automation is incapable of handling will be dealt with by human operators. AI is envisioned to have a central role in RCCs, with the aim of reducing the workload for operators responsible for monitoring and controlling autonomous vessels, as well as improve their decision-making capabilities. In general, AI should aid the human operator in achieving high level situational awareness. However, it has been shown that increased levels of automation can have negative effects e.g. decision biasing (Sarter and Schroeder, 2001). Furthermore, complex systems that are hard to understand can lead to out-of-the-loop performance, as operators are unable to identify the problem (Endsley, 2017).

Rødseth et al. (2022) discussed the use of an operational envelope as a tool to aid the human-machine interface in the context of autonomous ships. In cases where the risk is deemed too high for the automation to handle, a hand-over from the automation to the human operator must take place, or fail to safe for a minimum risk condition. The automation must, therefore, have a high degree of reliability within a constrained domain, as well as the ability to evaluate the risk associated with various situations. In situations where the risk is above a given threshold, the operator will be involved. High risk situations can, for instance, involve ship encounter situations or high traffic congestion. Furthermore, this can be extended to situations with high ML-model uncertainty, i.e. situations in which the ML-models may perform poorly. The operational envelope will define what the ship automation is expected to handle in this context.

Due to the close integration of human operators with AI-based systems, research has emerged on the topic of human-AI teaming (National Academies of Sciences and Medicine, 2022). National Academies of Sciences and Medicine (2022) identified, among other things, that explainable AI would play an essential role in human-AI teaming.

Endsley (2017) outlined transparency, understandability and predictability of the automation system as important factors in providing automation trust. ML-models are often referred to as "black-boxes", and are inherently lacking in many of these aspects. However, there has recently been an increased focus on trustworthy and explainable AI as an aid to humans interacting with AI-based systems. Through trustworthy AI, the user can be informed as to the limitations of the model, e.g. in cases of high model uncertainty. In such cases, the AI model is not capable of handling the situation, and human intervention is required. Explainable AI on the other hand, strives to provide the user with the reasoning behind a model prediction. If leveraged properly, such methods can assist RCC operators in overcoming many of the limitations of using AI in autonomous ships. A system involving human operators that can handle high risk situations, e.g. high AI-model uncertainty, should ease the approvability of AI-based autonomous ship systems.

5.2. *Explainable AI*

As mentioned, ML-models are often viewed as ”black boxes”. In this sense, the underlying causality associated with the prediction model is unknown, i.e. it is a ”black box”, within which the user has little insight. This had led to scepticism among end-users of AI-based systems, as well as degradation of situational awareness. Providing confidence in such models in safety-critical applications e.g. ship navigation can, therefore, be challenging. Models must also be transparent and understandable enough to enhance operator situational awareness.

In 2017, the Defense Advanced Research Projects Agency (DARPA) launched an explainable AI program to create more understandable AI systems via explanations catered to humans (Gunning and Aha, 2019). DARPA defines explainable AI as *”AI systems that explain their rationale to a human user, characterize their strengths and weaknesses, and convey an understanding of how they will behave in the future”* (Gunning and Aha, 2019). In this study, we decompose this definition into components of trustworthy and explainable AI, where explainable AI focuses on explaining why a decision was made. Trustworthy AI, on the other hand, deals with characterizing model strengths and weaknesses.

One of the most relevant application areas of AI is supporting the SAS, specifically with respect to object detection and classification. These models leverage deep convolutional neural networks, that are too complex to directly infer the casualty, and hence the explanation of a model prediction. One method to explain the reasoning behind deep ANN predictions is guided backpropagation (Springenberg et al., 2015). By evaluating the output of the network, the prediction is propagated backwards through the network to the input layer. This will yield a weighting over the input image indicating what the network has focused on. The user can then evaluate if the model is behaving in a manner based on causality, or if it rather is basing its prediction on spurious correlations. Over time, operators can gain confidence in model predictions. In many cases, however, the explanations may be too complex for human operators. Other related methods e.g. gradient-weighted class activation mapping (Grad-CAM) (Selvaraju et al., 2017) allow for the visualization of a heat map to illustrate what the model is focusing on, in a more interpretable manner.

However, methods must be developed to relate model explanations to the operator’s domain knowledge, i.e. increase their interpretability. If implemented incorrectly, the utilization of such techniques may compound the problem by providing additional information to the operator that they do not understand. Explainable AI techniques must, therefore, be designed to match the domain needs of the operator.

5.3. *Trustworthy AI*

Trustworthy AI attempts to evaluate model uncertainty. To aid in overcoming the challenges of approving AI-based models for use in autonomous shipping, an uncertainty-based system could be implemented. System approval would then be based on the ability of a human operator to take control when model uncertainty is high. As such, system approval would not require high model accuracy in all possible situations, but rather focus on their ability to predict their limitations. In most cases, i.e. where the input to the algorithm lies close to the mean of the training data, model uncertainty will be low. However, ML models are prone to inaccuracies when the input to the algorithm lies far from the distribution of the data upon which it was trained, or in regions of high variance in the training data. In such cases, model uncertainty will be high.

While the uncertainty of statistical models can often be calculated explicitly, ANN model uncertainty is not as easy to quantify. Nonetheless, research has emerged in recent years that attempts to address this issue (Abdar et al., 2021). Model uncertainty is generally considered to be comprised of aleatoric and epistemic uncertainty (Hüllermeier and Waegeman, 2021). Aleatoric uncertainty is due to the inherent randomness in the data set, i.e. noise in the data. Such uncertainty is, therefore, irreducible. Epistemic uncertainty, on the other hand, relates to uncertainty in model parameters due to a lack of information. It is, therefore, considered to be reducible by increasing the amount of training data. In the case of object classification in ship navigation, this may be challenging due to limited available data sets compared to those commonly used for computer vision. A high degree of epistemic uncertainty indicates, therefore, that the model does not have sufficient training data to conduct a reliable prediction. Alerting an operator to such situations may increase confidence in such models, especially in cases with limited training data.

One approach to estimate model uncertainty is to utilize a Bayesian Neural Network (MacKay, 1992). In such networks, a distribution of model parameters is utilized. The most common technique utilized is Monte Carlo (MC) dropout (Kendall and Gal, 2017). During inference, MC dropout will run multiple forward passes of the network, dropping random weights for each pass. If the uncertainty is low, the distribution of resultant predictions should be tightly bounded. A broad distribution, however, indicates greater uncertainty of the model. If leveraged properly, models can be made self-aware of their limitations by identifying situations with high uncertainty. In such high risk situations, a handover to an RCC operator should be conducted in a timely and effective manner, as the situation is outside of the operational envelope of the AI-based system.

6. Conclusion

The integration of AI-based functions into autonomous ship systems provides many challenges to approval processes. Performance can not be guaranteed in all possible situations, as AI models may perform in a sub-optimal manner in unforeseen situations. As such, this study found that AI should be utilized in a constrained domain, where the most relevant functions relate to object detection and classification. The study further suggests that humans can be utilized to aid in the approval process of such systems, by taking control in high-risk situations. Such situations may include complex navigation, or instances in which the uncertainty of an AI-based SAS is high. In this manner, AI models can be self-aware of their limitations, and allow for a safe handover to human operators when necessary. To facilitate such functions, recent developments within explainable and trustworthy AI may provide a solution, and it is recommended that approval processes take these into consideration. Nonetheless, such processes should involve extensive testing to ensure the robustness of AI-models to identify their limitations and facilitate a safe handover to a human operator.

Acknowledgement

The work presented in this text has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 815012 (AUTOSHIP) and No 859992 (AEGIS).

References

Abdar, M., F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya, et al. (2021). A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion 76*, 243–297.

Al Ridhawi, I., S. Otoum, M. Aloqaily, and A. Boukerche (2020). Generalizing ai: Challenges and opportunities for plug and play ai solutions. *IEEE Network 35*(1), 372–379.

BV, B. V. (2019). Ni 641: Guidelines for autonomous shipping.

DNV (2021a). DNV-CG-0264 Class Guideline - Autonomous and remotely operated ships.

DNV (2021b). DNV-RP-0510 Framework for assurance of data-driven algorithms and models.

Elshawi, R., M. Maher, and S. Sakr (2019). Automated machine learning: State-of-the-art and open challenges. *arXiv preprint arXiv:1906.02287*.

Endsley, M. R. (2017). From here to autonomy: lessons learned from human–automation research. *Human factors 59*(1), 5–27.

Feng, D., C. Haase-Schütz, L. Rosenbaum, H. Hertlein, C. Glaeser, F. Timm, W. Wiesbeck, and K. Dietmayer (2020). Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges. *IEEE Transactions on Intelligent Transportation Systems 22*(3), 1341–1360.

Gunning, D. and D. W. Aha (2019, jun). DARPA's explainable artificial intelligence program. *AI Magazine 40*(2), 44–58.

Gupta, K. M., D. W. Aha, and P. Moore (2009). Case-based collective inference for maritime object classification. In *International Conference on Case-Based Reasoning*, pp. 434–449. Springer.

Haugeland, J. (1989). *Artificial intelligence: The very idea*. MIT press.

Huang, Y., L. Chen, P. Chen, R. R. Negenborn, and P. van Gelder (2020). Ship collision avoidance methods: State-of-the-art. *Safety Science 121*, 451–473.

Hüllermeier, E. and W. Waegeman (2021). Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning 110*(3), 457–506.

IMO (1972). Convention on the International Regulations for Preventing Collisions at Sea, (COLREGs) - as amended.

IMO (2013). MSC.1/Circ.1455 Guidelines for the approval of alternatives and equivalents as provided for in various IMO instruments.

Kendall, A. and Y. Gal (2017). What uncertainties do we need in bayesian deep learning for computer vision? *Advances in neural information processing systems 30*.

Koopman, P. and M. Wagner (2017). Autonomous vehicle safety: An interdisciplinary challenge. *IEEE Intelligent Transportation Systems Magazine 9*(1), 90–96.

MacKay, D. J. (1992). A practical bayesian framework for backpropagation networks. *Neural computation 4*(3), 448–472.

Moosbauer, S., D. Konig, J. Jakel, and M. Teutsch (2019). A benchmark for deep learning based object detection in maritime environments. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 0–0.

National Academies of Sciences, E. and Medicine (2022). *Human-AI Teaming: State-of-the-Art and Research Needs*. Washington, DC: The National Academies Press.

Norwegian Maritime Authority (2020). Circular RSV 12-2020, Guidelines in connection with the construction or installation of automated functionality, with the intention of being able to perform uncrewed or partially uncrewed operation, In Norwegian.

Ravindran, R., M. J. Santora, and M. M. Jamali (2020). Multi-object detection and tracking, based on dnn, for autonomous vehicles: A review. *IEEE Sensors Journal 21*(5), 5668–5677.

Rødseth, Ø. J., L. A. Lien Wennersberg, and H. Nordahl (2022). Towards approval of autonomous ship systems by their operational envelope. *Journal of Marine Science and Technology 27*(1), 67–76.

Rødseth, Ø. J., L. A. L. Wennersberg, and H. Nordahl (2021). Improving safety of interactions between conventional and autonomous ships. *Proceedings of First International Conference on the Stability and Safety of Ships and Ocean Vehicles (STAB&S 2021)*.

Sarter, N. B. and B. Schroeder (2001, sep). Supporting decision making and action selection under time pressure and uncertainty: The case of in-flight icing. *Human Factors 43*(4), 573–583.

Schelter, S., F. Biessmann, T. Januschowski, D. Salinas, S. Seufert, and G. Szarvas (2018). On challenges in machine learning model management.

Selvaraju, R. R., M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Proceedings of the IEEE International Conference on Computer Vision 2017-Octob*, 618–626.

Shin, H.-C., K.-I. Lee, and C.-E. Lee (2020). Data augmentation method of object detection for deep learning in maritime image. In *2020 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pp. 463–466. IEEE.

Springenberg, J. T., A. Dosovitskiy, T. Brox, and M. Riedmiller (2015, dec). Striving for simplicity: The all convolutional net. *3rd International Conference on Learning Representations, ICLR 2015 - Workshop Track Proceedings*.

Thombre, S., Z. Zhao, H. Ramm-schmidt, J. Garcia, T. Malkamaki, S. Nikolskiy, T. Hammarberg, H. Nuortie, M. Bhuiyan, S. Sarkka, and V. Lehtola (2020, September). Sensors and ai techniques for situational awareness in autonomous ships: A review. *IEEE transactions on intelligent transportation systems*, 1–20.

Williams, A. P. and P. D. Scharre (2015). *Autonomous systems: Issues for defence policymakers*. NATO Allied Command Transformation.

Zhang, X., C. Wang, L. Jiang, L. An, and R. Yang (2021). Collision-avoidance navigation systems for maritime autonomous surface ships: A state of the art survey. *Ocean Engineering 235*, 109380.