Fraunhofer
CML

Fraunhofer Center for Maritime
Logistics and Services CML

BSH

BUNDESAMT FÜR
SEESCHIFFFAHRT
UND
HYDROGRAPHIE

VerifAI

Study on Objective-Based Standardization in the Verification and Certification of Intelligent Decision-Making Devices On-Board Semi-Autonomous Surface Vehicles

Titelbild: Fraunhofer CML/SHansche – istockphoto.com

Fraunhofer
CML

BUNDESAMT FÜR
SEESCHIFFFAHRT
UND
HYDROGRAPHIE

VerifAI

Study on Objective-Based Standardization in the Verification and Certification of Intelligent Decision-Making Devices On-Board Semi-Autonomous Surface Vehicles

DOI: 10.24406/publica-2899

Fraunhofer-Center for Maritime Logistics und Services

Paul Koch, M.Sc.
Thomas Stach, M.Sc.
Manfred Constapel, M.Sc.
Hans-Christoph Burmeister, Dipl.-Wirtsch.-Ing. Univ.

Contents

Index of illustrations

Index of tables

List of abbreviations

AI	Artificial Intelligence
AIS	Automatic Identification System
ANN	Artificial Neural Network
BSH	Bundesamt für Schifffahrt und Hydrographie (German Federal Maritime and Hydrographic Agency)
CI	Computational Intelligence
CNN	Convolutional Neural Network
COLREGs	Convention on the International Regulations for Preventing Collisions at Sea, 1972
DIN	Deutsches Institut für Normung (German Institute for Standardisation)
DL	Deep Learning
DVO	Durchführungsverordnung (Implementing Regulation)
EU	European Union
GNSS	Global Navigation Satellite System
IMO	International Maritime Organization
IMU	Inertial Measurement Unit
IPO	Input, processing and output
LSTM	Long Short-term Memory
MED	Marine Equipment Directive
MASS	Maritime Autonomous Surface Ships
ML	Machine Learning
MMSI	Maritime Mobile Service Identity
MRU	Motion Reference Unit
MS	Milestone
NMEA	National Marine Electronics Association
RADAR	Radio Detection and Ranging
RGB	Red, green and blue
sAI	Symbolic Artificial Intelligence
SOLAS	International Convention for the Safety of Life at Sea, 1974
WP	Work package

1. Introduction

1.1. Study Classification

Digital transformation is increasingly being adopted in the maritime sector. Partial and full automation of maritime processes has expanded significantly in recent years. Alongside classic approaches to rule-based control, new technologies rely on artificial intelligence (AI) to recognize situations affecting the navigation or operational safety of seagoing vessels proactively and in good time, and to react appropriately. Using AI might help ease the growing pressure on navigation officers resulting from increasingly busy maritime traffic and a lack of personnel (Brooks & Greenberg, 2022; Minter, 2021), and improve operational safety (Daranda & Dzemyda, 2020; Yoshida et al., 2021). Although interest in the benefits from adopting AI-driven systems in this sector is growing, a number of challenges remain to be overcome, particularly when it comes to adoption. This study will develop a guideline for verifiability, technology assessment and certification of these systems.

AI-based systems can imitate decision-making processes and make rational or rule-based decisions without explicit definitions for individual steps. They use a variety of mechanisms to interpret large amounts of data in real time with no human intervention, draw conclusions and operate independently (Norvig & Russell, 2021).

This study focuses on verifying AI-based systems that were also developed using machine learning methods. Such models come with a number of challenges that play a key role when designing verification and certification processes:

- Generalization of the operational domain
- Data quality management in design and validation processes
- Novel and complex model architectures

For the adaptation of verification and certification processes, this means on the one hand that systems are considered whose decisions are not readily comprehensible in every instance. On the other hand, due to their unpredictably growing diversity and multitude, the systems must be considered generically rather than individually in order to keep the effort for verification and certification processes within a feasible and sustainable framework.

1.2. Objectives and Requirements

The aim of this study is to develop a Verification Guideline for AI-based automated decision-making systems for maritime applications, as well as a Safety Guideline for manufacturers of AI-based maritime technologies. The Verification Guideline demonstrates how the German Federal Maritime and Hydrographic Agency (BSH) can verify for adequate safety and proper functioning. The Safety Guideline assists manufacturers in taking into account relevant aspects relating to safety during the development process, as well as help design an AI-based system that can be verified. The Verification and Safety Guidelines are designed to complement each other.

The study is divided into four main areas, which were worked on during the project:

- Review and evaluation of the current certification system (section 3) in terms of its suitability for verifying different types of AI-based systems. Initial results identify inadequacies in the current verification and certification system. This gives an idea of what is needed in the resulting work packages, and points of reference to existing processes can be generated.

- Assessing what information autonomous systems need, via a targeted market analysis of AI-based systems (section 4). The results will feed into developing the Verification Guideline for industrial systems, so that systems can move forward rapidly to the verification process.
- Developing a Verification Guideline (sections 5 and 6) which sets out the various steps required when verifying an AI-based system. The Verification Guideline sets out recommended procedures and defines the skills and processes required for certification.
- Developing a Safety Guideline (section 7) which ensures AI-based systems can be verified. The Safety Guideline also includes information that must be considered when designing an AI-based system in order to ensure operational safety.

1.3. Study Plan and Method

A period of 12 months was available to complete the study efficiently. The study itself was split into two work packages (WP) sections. The WP sections were further split into individual work packages. The time allocated to each WP is listed in the schedule Figure 1. The milestones (MS) are especially important since progress on the project is measured against these and, also, presented separately.

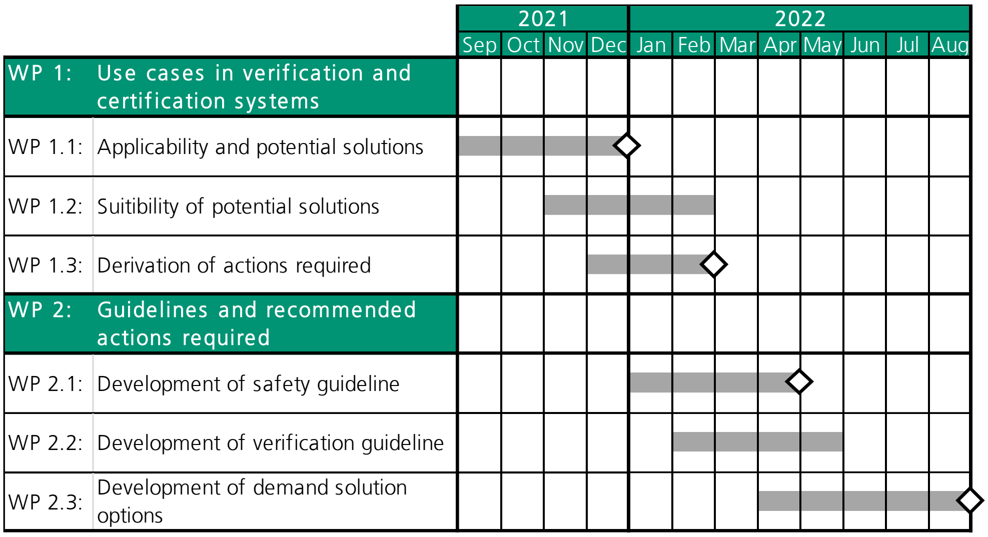


Figure 1: Schedule showing work package (WP) allocation for working on the study. The diamonds in this diagram represent milestones (MS).

In WP 1 (see "WP 1: use cases in verification and certification" in Figure 1), use cases were devised which were used to assess the suitability of the current verification and certification system and decide the need for any action. In WP 2 (see "WP 2: Verification Guidelines and recommendations for action" in Figure 1), a verification and Safety Guideline was devised as a recommendation for points of action. Each WP is assigned two MS. The four MS in all are listed in Table 1 and marked with diamonds in the schedule in Figure 1.

Table 1: Milestones of the projects.

#	Milestone description	Month
MS 1	Limits of the usefulness of the verification and certification system for the use cases to have been determined and documented.	4
MS 2	The suitability of alternative verification and certification scenarios in terms of ensuring operational safety to have been analyzed.	6
MS 3	A Safety Guideline to ensure operational safety using the findings from WP 1 to have been created.	8
MS 4	Options in terms of solutions to the requirements and recommendations for action based on the current state of science and technology to be established.	12

Beyond this organizational framework, there were regular meetings between the BSH and Fraunhofer CML. A fortnightly cycle was initially set for these meetings, which was expanded to a 3-week cycle once MS 2 was achieved. The final presentation was held on September 15, 2022, and the study was presented to the BSH the following month.

1.4. Study Structure

The study will examine the concept of "intelligent decision-making devices" with specific reference to "semi-autonomous surface ships" in terms of verification and certification capability. Assignments were worked on against the background of the defined WP and PS as set out in Figure 2.

The study is structured as follows. Starting with section 2 there is an introduction to essential terms and definitions. Section 3 begins with a review of the existing certification system and an examination of procedures currently used to assess conformity in the maritime sector. Building on the elements of the conformity assessment procedures considered, the limitations of these procedures when it comes to auditing AI-based systems are set out and explained in detail.

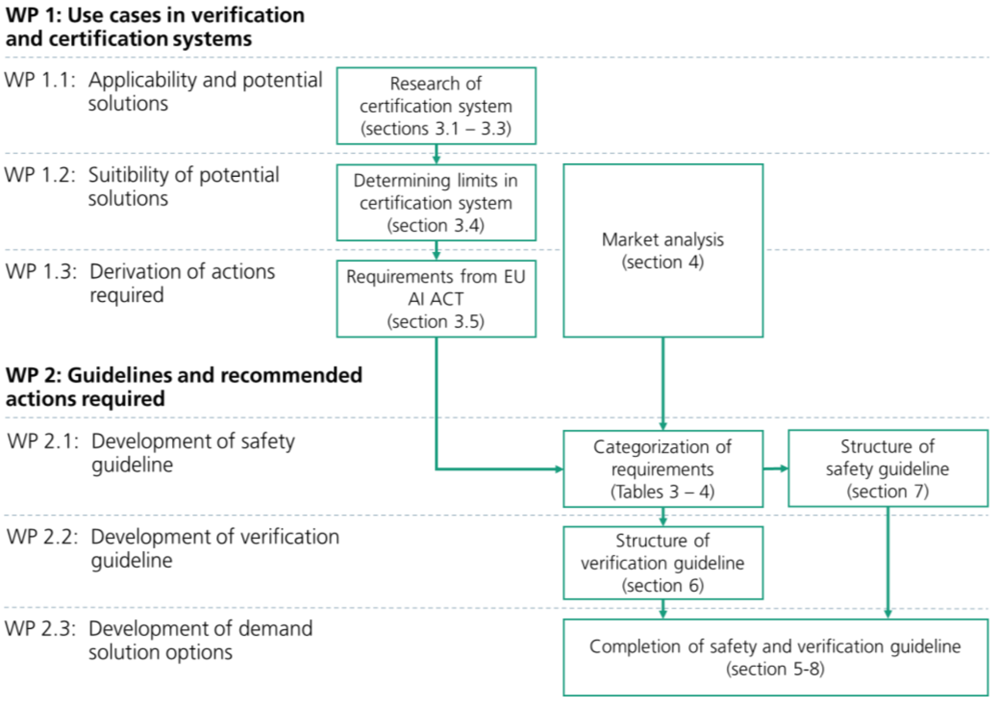

WP 1: Use cases in verification and certification systems
WP 1.1: Applicability and potential solutions
Research of certification system (sections 3.1 – 3.3)
WP 1.2: Suitibility of potential solutions
Determining limits in certification system (section 3.4)
Market analysis (section 4)
WP 1.3: Derivation of actions required
Requirements from EU AI ACT (section 3.5)
WP 2: Guidelines and recommended actions required
WP 2.1: Development of safety guideline
Categorization of requirements (Tables 3 – 4)
Structure of safety guideline (section 7)
WP 2.2: Development of verification guideline
Structure of verification guideline (section 6)
WP 2.3: Development of demand solution options
Completion of safety and verification guideline (section 5-8)

Figure 2: Plan of work for the study and sequence of tasks to be performed.

In section 4, in combination with the requirements identified and a market analysis carried out as part of the study, relevant information requirements are identified and categorized depending on the source of the data. For this, maritime technology involving AI that is already available but cannot yet be audited is examined and summarized in a market analysis. The results of this market analysis represent a key part of the study since they form the basis for designing the verification and Safety Guidelines.

Based on technical principles and the points of action identified, the verification and Safety Guideline, as well as how to integrate them, are introduced in section 5. A Verification Guideline intended for the BSH is then presented in section 6. The Verification Guideline consists of a sequence of audit steps which the BSH can use to verify AI-based systems taking a model-agnostic approach. In section 7, a safety concept intended for manufacturers is proposed. This sets out how manufacturers can prepare their system for verification so that it is auditable and has a good chance of being certified.

Sections 5 to 7 set out examples of how to apply the individual steps in the Verification and Safety Guidelines using an fictious AI-based product (see introduction in section 4.3). These examples are in turquoise boxes like this.

The study ends at section 8 with a summary and recommended points of action identified for the BSH.

2. Introduction to Terms and Definitions

The subject of artificial intelligence is finding its way into many different industries — including the maritime sector. The terms artificial intelligence and autonomous shipping as well as other related terms are often defined or understood differently. A common understanding of terms and concepts is needed for basic comprehension, as well as an explanation of how these are linked.

2.1. Artificial Intelligence as a Decision-making Device

With the development of the first computers and their ever-increasing computing power, the subject of artificial intelligence not only drew the attention of science, but increasingly many areas of industry too. This study considers artificial intelligence systems as decision-making devices for making safety-critical decisions on board surface ships with semi-autonomous capability.

Even the experts disagree on when a system can be termed AI (Norvig & Russell, 2021). One idea for a definition of AI might, for instance, be to evaluate how rational and correct, or how human-like the decisions made by a system are. This way of looking at a system might be approached both internally, meaning with regard to a series of internal conclusions, and externally, meaning in terms of external performance or results.

In this subsection, the terms machine learning and deep learning are classed as subsets of artificial intelligence, as depicted in simplified form in Figure 3.

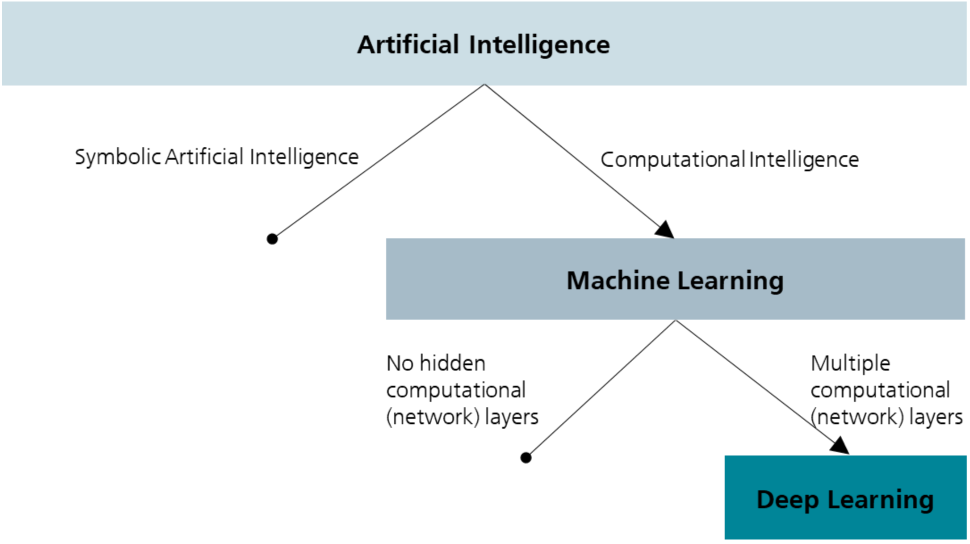

Artificial Intelligence
Symbolic Artificial Intelligence
Computational Intelligence
Machine Learning
No hidden computational (network) layers
Multiple computational (network) layers
Deep Learning

Figure 3: Relationship between artificial intelligence (AI), machine learning (ML) and deep learning (DL).

Basically, AI approaches can refer to either of the following techniques: Symbolic Artificial Intelligence (sAI) or Computational Intelligence (CI) (Flasiński, 2016). An sAI model is characterized by the fact that it can be defined using explicit terms. This means that knowledge is represented symbolically and "thought" processes in the model are defined as formal operations. Explicitly formulated decision trees or expert systems can be cited as examples here. In a CI model, however, information is generally represented numerically. This means that "thought" processes are mainly carried out in the form of numerical calculations and knowledge is not necessarily stored as explicit terms. One example is neural networks, where knowledge is stored in the form of a network with numerically weighted nodes. A CI model is one which was created using an algorithm based on an existing data set. The process of generating such a model in general is known as machine learning (ML) and the creation and subsequent improvement of the model more specifically is known as training.

2.1.1. Machine Learning

Two essential elements in ML are the data set and the algorithmic approach. The data set represents the "experience" which is used to train the AI model. Training then ensues using algorithms. Data set, the algorithmic approach and the application being pursued are interrelated and cannot be selected independently of one other.

The data sets and data used in ML can be very different. It can therefore be helpful to characterize these. Modality of data sets can vary and include data types such as image, text, audio, sequential or tabular data. Besides these and other possible data types, the statistical distribution of the data also plays a key role. For an AI model to function effectively in the real world, the statistical distribution of the data set used for training must reflect the expected statistical distribution from the real-world application. The statistical distribution of the data can be characterized using descriptive statistical methods (Navlani et al., 2021).

Depending on the intended application and the available data set, different ML paradigms are suitable. An ML paradigm can usually be classified as one of the following three basic learning paradigms (Burkov, 2019):

- Supervised Learning: Training using a data set where the target output values for corresponding input values are stored.
- Unsupervised Learning: Training using a data set with no predefined target output values.
- Reinforcement Learning: Training which uses trial and error to learn from decisions.

Once training a model is complete, then that model can be considered as frozen. When this happens, the model's performance is deterministic and therefore reproducible. Because the expected performance does not change over time.

Besides being classified into learning paradigms, ML paradigms can also be classed in terms of the task to be solved. There is a fundamental relationship between the learning paradigm selected and the types of tasks which can be solved. This is explained below in Figure 4. There are essentially three types of task as follows (Burkov, 2019):

- Classification: With classification, incoming data is assigned output values that, if the output is valid, correspond to the expected target values[1]. A target value is an element of a set from different classes. Classes represent discrete values and can, for example, represent character strings, such as texts, as a data type.
- Regression: As with classification, in regression target values are assigned to the incoming data. It differs from classification in that these are target values from a continuous space — real-valued functions, for examples.

[1] Discrete target values, especially those relating to classification problems, are usually referred to as "labels" in technical literature. In this study, the more general term "target values" is used.

- Clustering: The fundamental difference between clustering and classification or regression is that there are no corresponding target values for the input values. Clustering organizes similar data objects into groups based on a data set. New input values can be grouped depending on the clustering criteria.

Artificial Neural Networks (ANN) are a special case of machine learning techniques. The name stems from the fact that their structure is based on the neural networks in the human brain (Flasiński, 2016). ANNs essentially have an input layer, one or more hidden computation histories, and an output layer. Neurons between adjacent layers can be connected to each other as in a network. ANNs cannot be clearly assigned to a learning paradigm or task type. Depending on technique and structure, an ANN can serve different applications, such as object detection or time sequence prediction.

Some notable examples of ML techniques are listed in Figure 4, sorted hierarchically in terms of underlying learning paradigms and task types. The diagram depicts just a selection of possible learning paradigms, task types and techniques. For a wider overview and more examples please see (Sarker, 2021).

One well-known classification technique is the decision tree (Burkov, 2019; Sarker, 2021). A variety of algorithms can be used to generate a decision tree. The model ultimately comprises a root node and several decision nodes. These represent branches, along which the route forward is evaluated using calculations, which then eventually leads to one of many leaf nodes. The output values are determined and output at the leaf nodes.

A support vector machine (SVM) is a machine learning approach suitable for both classification and regression problems. This approach involves organizing data points into groups using one or more separation hyperplanes (or in the linear, two-dimensional case using separation lines). The separation hyperplane is chosen where the distance to the data points is greatest. Classification is therefore made possible by partitioning the data space using the separation hyperplanes. Variations of the SVM approach allow the separation hyperplane to be used as a regression hyperplane (Burkov, 2019; Sarker, 2021).

One common regression method is linear regression (Burkov, 2019; Sarker, 2021). This involves drawing a straight line through the input values. The best-fit line can be optimized using the least squares method, for instance. Using the best-fit line, approximate output values can be read against input values.

A simple example of a clustering method is the k-means clustering method (Burkov, 2019; Sarker, 2021). With this method, a number of k clusters or data point groups, are created. Each group has a geometric center. New data points are assigned to a group using a measure of distance to calculate which geometric center of gravity from which group it is closest to.

DBSCAN is a clustering method which can be used for shipping traffic data (Riveiro et al., 2018). Unlike with k-means, with DBSCAN the number of clusters to be created is not specified in advance. Instead, it defines how close data points need to be and how many sufficiently close data points are required so they qualify as an independent cluster (Burkov, 2019). This approach means that not all data points can or need belong to a cluster.

In reinforcement learning (RL) methods, actions such as decisions are implemented in sequence by an agent (Sarker, 2021). Each action brings with it a reward or punishment, which forms the basis for adapting the extent to which the agent perceives its environment. Ideally, the agent adapts to its environment over time, so that the best-possible decisions are made. Reinforcement learning distinguishes between model-based and model-free approaches. With a model-based approach, the agent attempts to build a model of its environment by comparing its decision and the associated reward with its original prediction. It makes decisions taking its perception (prediction) of the environment into account. Agents in a model-free approach make decisions based purely on experience, without making any *a priori* predictions.

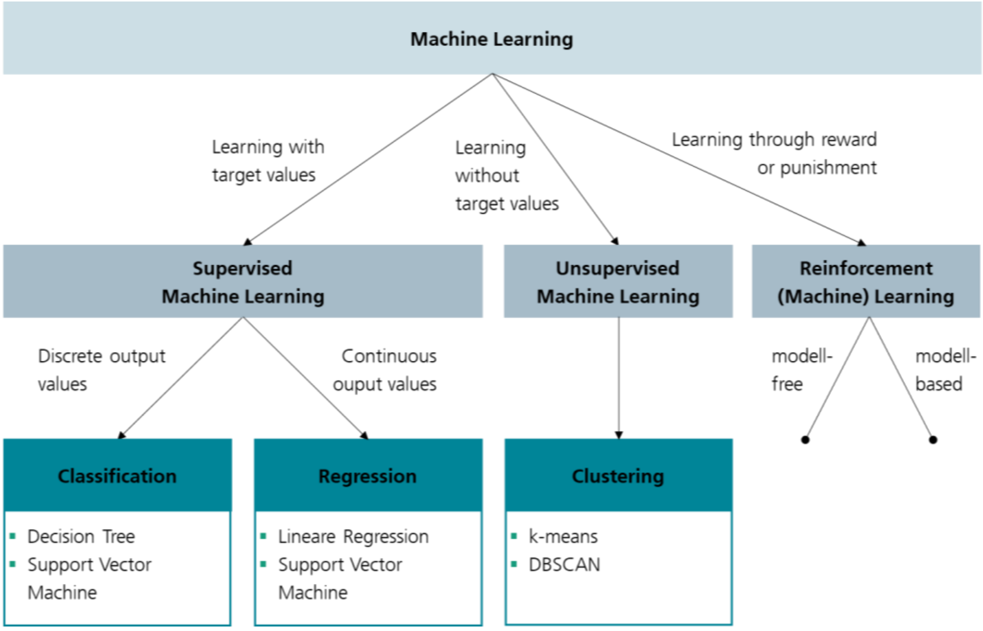


Figure 4: Examples of machine learning methods sorted by learning paradigm and task type.

Where multiple computing layers are used in a machine learning approach, this is referred to as Deep Learning (DL) (Norvig & Russell, 2021). Deep Learning methods are used, for example, in visual object recognition, where a convolutional neural network (CNN) is usually used as a model. If the model being trained using DL is a neural network, the model is also referred to as a deep neural network.

Both in the development phase when training a machine learning-based model and during the application phase it may produce less reliable output. A well-known, but solvable, problem is underfitting or overfitting (Burkov, 2019; Norvig & Russell, 2021). When a model delivers too many errors in its training data, in the form of wildly differing output values, this is called underfitting. Underfitting can result from data that is uninformative for the model or from choosing a model that is too simplistic. Figure 5 illustrates this using the example of linear regression, which is approximated at data points that roughly follow a quadratic function. With overfitting, on the other hand, the model is too biased towards the training data and fails to produce a correct output for input values that were not included in the training data. In Figure 5 this can be seen from the fact that the fit hits all the data points, but no longer follows the roughly quadratic curve of the underlying function. Overfitting can result from a model which is too complex (in this example, a function where the degree of approximation) or from too little training data. When it comes to overfitting, it is also said that the model has a high variance because it has a wide degree of spread.

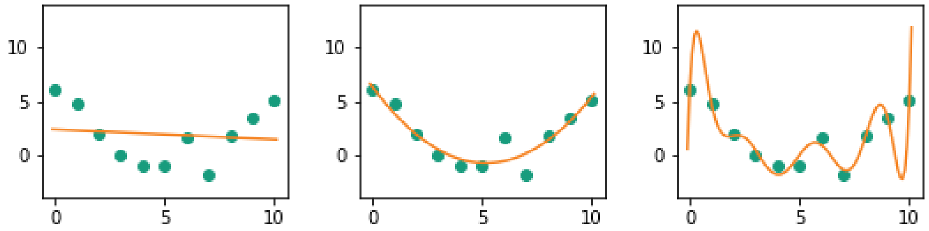


Figure 5: A quadratic distribution of points with outliers (green) and related adjustment functions (orange). Underfitting (left), good fit (middle) and overfitting (right).

The performance of AI-based systems that worked successfully during the development phase may deteriorate over time in the application phase. This is caused not so much by an intrinsic deterioration in the AI-based system itself (where the performance is considered frozen), but instead by changes in the environment in which the AI-based system is operating. This changing relationship between model and reality is known as concept drift (Žliobaitė, 2010). Drift can take different forms:

- Statistical distribution of data changes over time.
- The relationship between model input and output changes over time.
- The ground truth assumed at the outset, such as the designation or selection of target values, changes.

This is why it plays a major role in testing the reliability of (frozen) AI-based systems even after they have been developed.

2.1.2. Transparency of decisions

The transparency of decisions made by AI-based models depends heavily on the ML approach chosen for the model. The ML approach determines whether an artificial decision-making mechanism consists of a deep neural network or decision tree, and so whether the model is less or more transparent. The transparency of a system can be defined in terms of how easily it can be interpreted or explained. Ease of interpretation or explanation are different terms in the context of artificial intelligence (Norvig & Russell, 2021).

Models with high transparency are those whose internal structure makes it intuitively clear why the model reaches the decisions it makes. One example of this is a decision tree, because its simple if-then structure reflects the route to a decision in a way which humans find easy to comprehend. The decision-making process can therefore be understood without an external system. This is depicted schematically in Figure 6 on the left.

By contrast, a deep neural network is not easy to interpret and therefore not transparent (see Figure 6, on the right). Looking at its structure, a large number of neurons can be seen (see pale blue dots in Figure 6, on the right), which are built up in parallel in computing layers connected in series (see rows of pale blue dots in Figure 6, on the right). Each neuron consists of an activation function and a weighting between 0 and 1 (as opposed to an if-then structure). The model and its decision-making are therefore inaccessible to human understanding. In this case the model is called a black box. Ease of explanation can be introduced using an external module. For instance, in the example of visual ship recognition, a module might be used that marks features it considers typical of an object of the type "ship". The research field that deals with the ease of explanation of AI is referred to as Explainable AI (Samek & Müller, 2019).

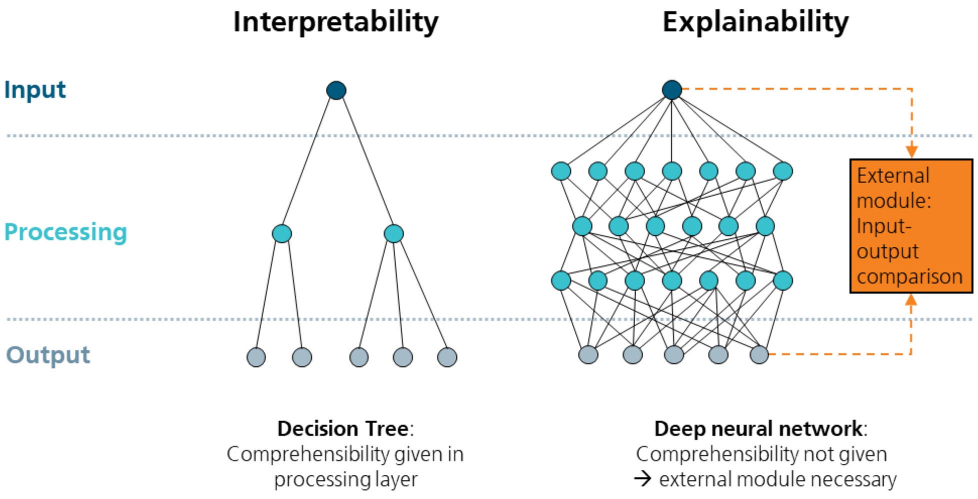

Interpretability
Explainability
Input
Processing
Output
External module: Input-output comparison
Decision Tree: Comprehensibility given in processing layer
Deep neural network: Comprehensibility not given → external module necessary

Figure 6: Distinction between ease of interpretation and explanation in AI models. This example shows a decision tree (left) and a deep neural network (right). In order to make the neural network transparent, an external additional module is required to compare inputs/outputs (orange box).

Weller introduces the concept of transparency, which meaning depends on which group of people encounter the AI-based system under consideration (Weller, 2019). Transparency means something different to manufacturers than to regulatory bodies. The former is interested, for instance, in how the neurons within a CNN contribute to the classification of certain objects. The interest of the latter, however, should focus particularly on checking the accuracy of decisions made by an AI-based system. Put simply, manufacturers are attempting to answer the question "how", and regulatory bodies are attempting to answer the question "whether". This means that Explainable AI, which deals with the question of "how", is not considered or pursued as a possible solution in this study.

2.2. Semi-autonomous Surface Ships

Maritime autonomous surface ships, known as MASS, are ships which, to a varying degree, can operate independent of human interaction. Operation includes not only the sub-tasks of ship operation such as monitoring the engines, situation awareness or navigating, but also the full task of safe ship navigation.

The term autonomous does not describe self-determination by the system in the true sense of the word, but rather the automation of operating processes (Etzkorn, 2022).

Studies by the International Maritime Organization (IMO) have determined that semi- and fully autonomous surface ships can be divided into four degrees of autonomy (IMO, 2022), as shown in Table 2. The point of this subdivision is to categorize developments in the MASS sector and identify challenges in the certification and operation of MASS.

Table 2: Degrees of autonomy per IMO studies as part of the IMO Scoping Exercise.

Degree of autonomy	Meaning
1	*Ship with automated processes and decision support:* Seafarers are present on board to operate and control systems and functions. Some processes are automated and unsupervised, but humans can intervene at any time.
2	*Remotely controlled ship with seafarers on board:* The ship is controlled and operated from a separate location. Seafarers are on board to take control and monitor the systems and functions.
3	*Remotely controlled ship with seafarers on board:* The ship is controlled and operated from a separate location. There are no seafarers on board.
4	*Fully autonomous ship:* The ship's operating system can make its own decisions and determine its own actions.

This study focuses on semi-autonomous surface ships, meaning ships equipped with autonomous systems or where certain processes are replaced by automatic decision-making systems. Remotely controlled systems were not included in this study. Consequently, this study focuses on ships and ship technologies for autonomy levels 1 and 4 per the IMO.

The study looks at individual systems with domain-specific tasks, so simplifying the design of verification processes. As part of the study, complex AI-based systems are broken down into their component parts and verified separately. The verification and Safety Guideline further aims to develop broadly applicable processes, regardless of the degree of autonomy, which can be transferred to individual AI-based systems as well as to a larger collection of AI-based systems.

3. Verification and Certification Systems

According to the International Convention for the Safety of Life at Sea (SOLAS), placing new kinds of component parts for ships on the market (marine equipment) requires the function of the device itself to be tested and certified, along with the manufacturing process and operation of the equipment on board the ship. Thorough verification is required to ensure operational safety of these systems, especially when it concerns approving systems intended to autonomize processes on board ships. This section explains the processes needed for verification and approving system parts and reveals the limitations of existing procedures.

3.1. EU conformity assessment procedure

Verifying the safety of marine equipment in the European Union (EU) is carried out by notified bodies in accordance with the Marine Equipment Directive (MED) through a conformity assessment procedure. Notified bodies are institutions accredited by national authorities and mandated to carry out verification procedures. Ensuring conformity of products to be placed on the market in the EU covers design, construction and performance. The EU has set out the conformity assessment procedure and the various conformity assessment modules and options as part of the Marine Equipment Regulations (Europäisches Parlament und Rat der Europäischen Union, 2014).

Broadly speaking, the conformity assessment procedure can be divided into two verification options depending on the manufacture method, which can sometimes be subdivided into more advanced modules. The modules shown in Figure 7 are explained in the following sections and considered in terms of their objectives.

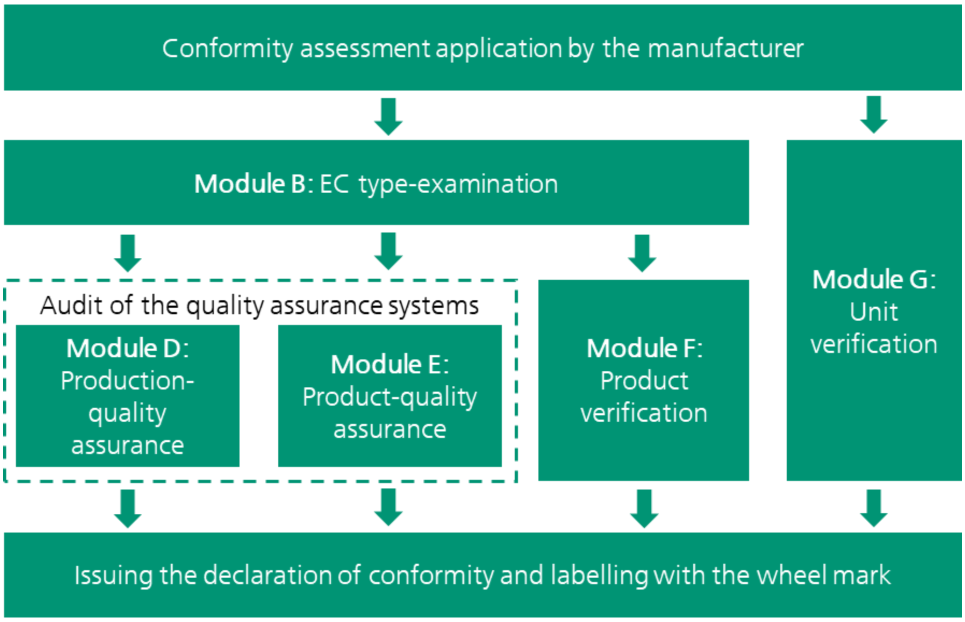

Conformity assessment application by the manufacturer
Module B: EC type-examination
Audit of the quality assurance systems
Module D: Production-quality assurance
Module E: Product-quality assurance
Module F: Product verification
Module G: Unit verification
Issuing the declaration of conformity and labelling with the wheel mark

Figure 7: Conformity assessment procedure under the MED.

Where a product is manufactured in series or in mass, an EC type-examination (module B) (EC stands for European Community) is required, as well as one of the following conformity assessment modules, depending on the manufacturer and type of product (Europäisches Parlament und Rat der Europäischen Union, 2014):

- Production-quality assurance (module D)
- Product-quality assurance (module E)
- Product verification (module F)

Where a product is not manufactured in series or in mass, but either individually or in small quantities, safety and quality are inspected with an EC unit verification (module G) (Europäisches Parlament und Rat der Europäischen Union, 2014).

3.1.1. Module B – EC type-examination

The type examination examines the technical design of marine equipment for safety and conformity. The aim of this examination is to ensure the adequacy of the technical design, compliance of the sample with the technical documentation, and conformity with existing standards and guidelines.

The examination can be carried out on representative samples of either the end product or on one or more major parts of the product. The second option also requires an assessment of the adequacy of the technical design based on technical documentation and evidence (Europäisches Parlament und Rat der Europäischen Union, 2014).

3.1.2. Module D – Production-quality assurance

Module D examines the quality assurance system of the production process for conformity to international instruments and standards. The system to be inspected must ensure both conformity of the products with the design previously examined in Module B and the quality objectives applying to the products manufactured, throughout the entire production process. To ensure this, product quality is inspected before, during and after production (Europäisches Parlament und Rat der Europäischen Union, 2014).

3.1.3. Module E – Product-quality assurance

For conformity assessment module E, conformity is examined by inspecting the product’s quality assurance system. As in module D, the system to be inspected must ensure compliance of the products with the design examined in Module B. It must also ensure that the products achieve the product quality objectives at the end of the production process. To ensure this, the product quality is inspected once final production has been approved. (Europäisches Parlament und Rat der Europäischen Union, 2014).

3.1.4. Module F – Product verification

For conformity assessment module F, individual products are checked for conformity by the notified bodies following the production process. To do this, the manufacturer must first ensure that the products comply with the design checked in module B and that the requirements for the manufacturing process comply. In a further stage, the manufacturer can choose between a verification method based on verifying every product, and testing on a statistical basis. Where the latter is chosen, every batch produced must be consistent (Europäisches Parlament und Rat der Europäischen Union, 2014).

3.1.5. Module G – Unit verification

Where a product is manufactured individually or in small quantities, it is tested individually under module G as part of the conformity assessment procedure. In a first stage, the supporting technical documents must be examined for conformity with existing standards and directives, and design calculations and tests must be carried out. In a further stage, individual products are checked against previously defined requirements (Europäisches Parlament und Rat der Europäischen Union, 2014).

3.2. Implementing regulation

Besides the MED, which specifies the conformity assessment procedure, the European Commission has published the Implementing Regulation (Europäische Kommission, 2022). This specifies the various design, construction and performance requirements as well as the verification standards for marine equipment. The different equipment is classified and published in the following nine categories (Europäische Kommission, 2022):

1. Life-saving appliances
2. Marine pollution prevention equipment
3. Fire protection equipment
4. Navigation equipment
5. Radiocommunication equipment
6. Equipment required under the 1972 Collision Prevention Regulations (COLREGs)
7. Other safety equipment
8. SOLAS Chapter II-1 equipment
9. Equipment for which the set of standards for MED certification is not complete

If marine equipment can be classified under categories 1 to 8 and no standards and requirements required by the IMO are available or appropriate, it is classified under category 9. (Europäische Kommission, 2022).

3.3. National certification

National certification can be granted for marine equipment that is subject to certification under German legislation and for which there are no internationally harmonized requirements. This certification must be recognized by other European member states where the guaranteed level of safety complies with the rules of the respective nation (Bundesamt für Schifffahrt und Hydrographie, 2022).

3.4. Inadequacies in the identified standards and guidelines

In order to certify AI-based systems on ships using the EU conformity assessment procedure, it must first be established whether the various test elements do apply to these systems. To do this, the current Implementing Regulation and verification procedures are examined below for conformity with the requirements for AI-based systems.

The current version of the Implementing Regulation (Europäische Kommission, 2022), dating from 2021, does not include any procedures for verification of AI-based systems. There is no potential category for classifying such systems and no procedure by which AI-based systems can be incorporated. Registration under the Implementing Regulation calls for appropriate standards and specifications defining the necessary requirements for such a system. Verification standards for an EU conformity assessment procedure can normally be established by different organizations (Europäisches Parlament und Rat der Europäischen Union, 2014). The IMO is in particular responsible for developing international standard regulation of autonomous and semi-autonomous vessels (Danish Maritime Authority, 2018).

In order to lay the foundations of a regulatory framework for AI-based systems, the IMO has established a road map for developing one. Consequently, an international regulatory framework for MASS (maritime autonomous surface ships) is set to be introduced in 2028 (IMO, 2022). At the national level, an initial review of the DIN standards currently in force reveals that although there are isolated standards, specifications and draft standards (DIN-SPEC), these do not define the requirements for an AI-based system in a model-agnostic manner. After reviewing all the documentation evaluated as part of the study, it was found that existing DIN standards apply to specific individual cases and do not currently allow general certification.

Besides standards and specifications, suitable tools are required within a verification procedure for certification. An investigation into the EU conformity assessment procedure shows that the verification modules set out above cannot be used for certification because the verification methods and tools these use turn out to be unsuitable. AI-based systems are not covered by a type examination (module B) nor within the product quality and quality assurance procedures (modules D, E, F) or the currently defined methods for unit verification (module G) (Europäisches Parlament und Rat der Europäischen Union, 2014). This applies to all existing modules and relevant regulations. This means that AI-based systems based on CI models (see section 2.1) cannot be reviewed against existing verification procedures.

There is therefore an urgent need for action to develop dedicated processes which examine the performance of AI-based systems to ensure they are functioning correctly. In order to include new verification procedures, the existing conformity assessment procedure should be expanded to include an additional module specifically designed for AI-based systems, in order to be able to verify AI-based systems using existing and future procedures. One possible extension to the MED conformity assessment procedure is set out in Figure 8.

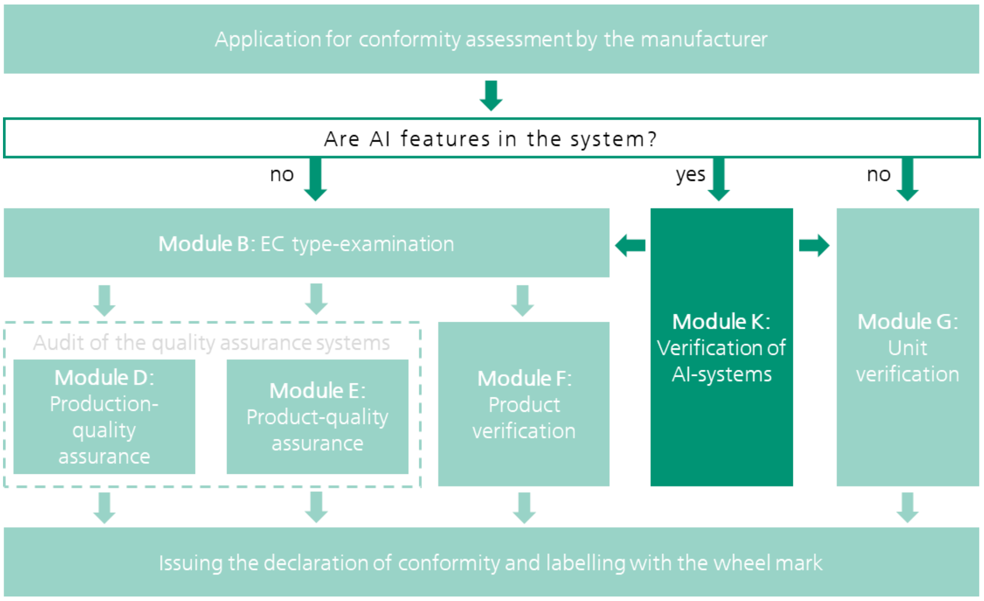

Application for conformity assessment by the manufacturer
Are AI features in the system?
no
yes
no
Module B: EC type-examination
Module K: Verification of AI-systems
Module G: Unit verification
Audit of the quality assurance systems
Module D: Production-quality assurance
Module E: Product-quality assurance
Module F: Product verification
Issuing the declaration of conformity and labelling with the wheel mark

Figure 8: Introduction of module K in the conformity assessment procedure under the MED.

The proposed conformity assessment module K closes the current regulatory gap for certification of AI-based systems specifically in the maritime sector. The recommended text of this conformity assessment module was devised as part of this study and can be found in the Verification Guideline in section 6.

Alongside this approach, the European Union is currently developing a more general proposal for the certification of AI-based systems. In order to anticipate the potential changes arising from adopting this proposed legal framework, this will be examined in more detail in the following subsection.

3.5. EU regulatory framework proposal on artificial intelligence

The proposal for a regulatory framework on artificial intelligence, also called the EU AI Act, was tabled by the European Commission in April 2021 (Europäische Kommission, 2021). The four-level risk-based approach distinguishes between unacceptable, high, low and minimal risk systems. The procedures proposed currently serve only as a draft procedure for a future regulatory framework and may not reflect all future requirements. It is reasonable to assume that the requirements already itemized will essentially remain the same and should be taken into account for an initial draft of conformity assessment procedures for semi-autonomous surface ships.

3.5.1. Risk-based classification

According to the proposed regulatory framework, AI-based systems with unacceptable risks could effectively be banned in the European Union (EU). This risk category would include practices which breach fundamental EU rights and values or have the potential to be manipulative and exploitative.

One level down from this are high-risk systems. If the regulatory framework is adopted, these would be subject to a conformity assessment procedure where the systems are examined for compliance with various criteria. The proposed regulatory framework for high-risk systems and their use cases are explained in more detail in the following subsection. Minimal or low risk AI-based systems would only be subject to a minimal transparency requirement (Europäische Kommission, 2021).

3.5.2. High-risk AI-based systems

One key proposal in the EU AI Act is the regulation of high-risk systems. This concerns AI-based systems that serve either as safety elements for products requiring prior conformity assessment or otherwise affect fundamental EU rights. Since the MED requires a conformity assessment for ship equipment, it can be assumed that it will be classified as a high-risk system (Europäische Kommission, 2021).

The conformity assessment procedures for marine equipment with integrated AI-based systems proposed in the EU AI Act as it currently stands would rely on two assessments: an assessment of quality management and an assessment of the technical documentation (Europäische Kommission, 2021).

Developers of high-risk AI-based systems would be required to set up a suitable quality management system. This would need to be able to prove as part of the audit process that it guarantees the quality and safety of the system over the entire life cycle. To ensure quality, certain minimum requirements for technical systems and documentation must be met. These include, for example, procedures covering data management, quality assurance, risk management or for reporting faults. In order to check the quality management system is being applied and maintained effectively, notified bodies might carry out regular audits (Europäische Kommission, 2021).

Besides the quality management system, developers would also have to provide evidence of full technical documentation for conformity in the EU. This would include a detailed description of all technical systems as well as all hardware and software elements used or which provide support over the life cycle. The technical documentation also requires a list of all harmonized standards and technical specifications applied (Europäische Kommission, 2021).

Generally speaking, the two assessment procedures should ensure that the following requirements are met (Europäische Kommission, 2021):

- A suitable risk management system that systematically identifies and analyzes risks.
- A data governance and data management system that ensures high, consistent data quality.
- Full technical documentation.
- Mandatory recording of all processes and events.
- Transparent provision of information to users.
- Constant human oversight.
- An appropriate level of accuracy, robustness and cybersecurity.

As the EU AI Act is only in draft form, it is possible that individual elements might change during the legislative process. In order to be prepared for the extensive impact of this legislation covering AI certification, this study assumes that the regulatory framework will be adopted within the next few years.

4. Market Analysis of AI-based Systems in a Maritime Setting

As already set out in the preceding sections, there is at present no suitable procedure for the audit of AI-based systems. As things stand, the safety of AI-based systems is therefore in the hands of the companies and their internal processes that are not publicly accessible. This current regulatory gap is becoming increasingly important as the digitalization and autonomization of shipping progresses. Thanks to the increasing number of technologies being placed on the market, the pressure on regulatory bodies and competent authorities is building. The adoption of autonomization can be seen by looking at the history of international patent applications in the MASS sector. Figure 9 shows the increasing volume of patent applications filed annually from 1990 to 2021. This shows patent applications that can be found using the combination of the two search terms “autonomous” and “ship”. The shape of the curve can be an indication that the number of AI-assisted products placed on the market each year will continue increase.

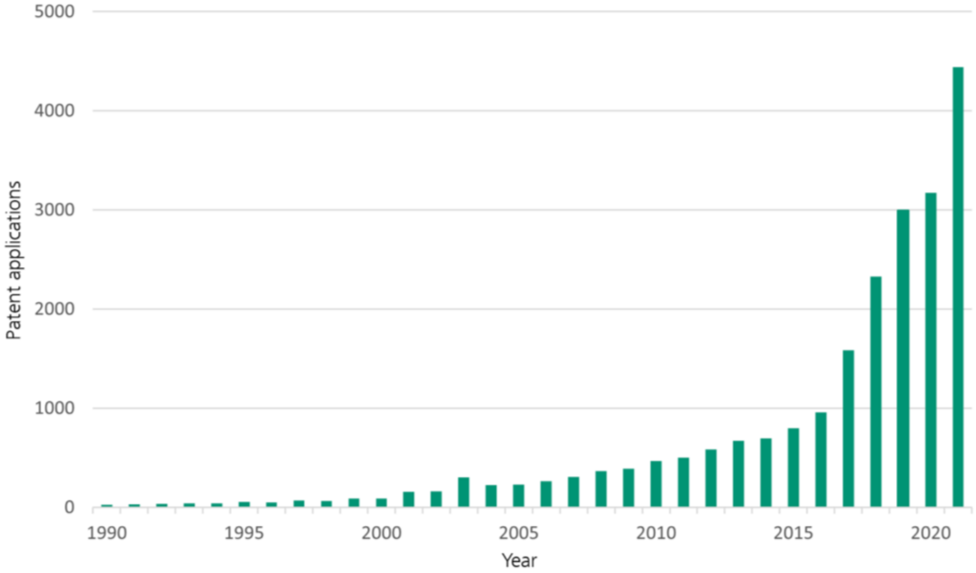


Figure 9: Patent applications relating to MASS, 1990 to 2021. Extracted from “Google Patents” (Alphabet Inc.).

With reference to the inadequacy of existing standardizations and certification procedures as well as the lack of national and international processes and standards identified in section 3.4, there is a need to establish appropriate verification and certification processes. Particularly on the assumption that such an assessment procedure guarantees the security protocols and mechanisms when implementing AI algorithms, the pressure to develop useful guidelines for auditing AI-based systems in the short term is growing.

In order to gain an understanding of which products require certification and what such certification might look like, existing MASS products with AI support are identified below and analyzed according to various criteria. The selection of criteria results from the range of sensors, which is listed in the context of MASS-related developments and publications and appears necessary for successful implementation. The market analysis was carried out with a focus on semi-autonomous surface systems with the aim of identifying the information needs of relevant systems and categorizing them. Once identified, information needs are used to research existing standardizations of the data used with the aim of standardizing these within the framework of the verification and Safety Guideline. The section ends by setting out a fictitious use case, which is used to exemplify the verification and Safety Guideline.

4.1. Review and Analysis of AI-assisted Products

The following market analysis examines various AI-assisted products currently on the market based on two key questions:

- Which data sources do the products use?
- What use cases do the products serve?

As part of the market analysis, 18 products from 16 different companies were reviewed. Of these systems used in autonomy, 17 are installed on the ships themselves and one on land. Depending on their purpose and use, the products reviewed differ in their degree of autonomous aspects. The products reviewed range from simple, camera-based berthing assistance to fully autonomous container ships. These are manufactured by small, medium-sized and large companies from across the world. The information gathered is based on the companies' product literature, which has been examined to address the key questions mentioned.

Which data sources the products use and which use cases they serve is detailed below. A detailed overview of the systems examined is provided in Appendix A.1. in Table 5 and Table 6.

4.1.1. Categorizing the Data Sources

The data sources identified for the systems examined can be classified based on the sensors used. Categorization according to sensor technology helps by focusing on the type and form of data that is captured and processed by the system.

Table 3 below summarizes all sensor systems identified and, where available, references to standardizations. To assess how useful these are for the development of maritime AI-based systems, they are analyzed with a focus on existing standardization by the IMO. This focused view allows an assessment of the extent to which existing sensors can essentially already be used for maritime applications, and whether certification in AI-based systems only needs to be sought with a view to the AI elements themselves.

Table 3: Sensor systems for data sources from AI-based systems and their communication standards identified in the market analysis. The abbreviations for these sensors are included in the list of abbreviations at the beginning.

Data source sensors	Performance requirements per IMO
AIS	(IMO, 2015; ITU, 2014)
GNSS	(IMO, 1995, 2001)
IMU/MRU	(IMO, 2017)
Infra-red and RGB[1] camera systems	No
LIDAR	No
RADAR	(IMO, 2004)
Depth gauges	(IMO, 1971, 1998)
Weather sensors	No

Of the sensors for which communication standards already exist, RADAR and AIS in particular are often mentioned by manufacturers as part of the products. RGB camera systems are used in all products reviewed, although they do not follow any standardization. In particular, the use of camera systems and the resulting image-based AI-based systems poses hurdles due to a lack of standardization in the maritime setting. AI-based systems which rely on such sensors should be subject to additional verification and certification compared to sensors with standardized information exchange. The background to this is that standardization of information exchange can actually provide information about the expected input and output data.

4.1.2. Use Case Analysis

In a first stage, evaluation of the use cases resulted in categorization into four groups into which AI-based systems can be classified: identifying objects, predicting road user behavior, route planning and situation awareness. Categorization helps to classify product groups in terms of their basic data and to extract relevant information requirements.

Object recognition using camera-based data is one application used in various forms in all products reviewed. The objects can be obstacles, other ships or the coast. These objects are sometimes identified using AIS data and output for decision support. Object recognition can also be useful as a system to help with berthing or unberthing, by providing the ship's command with relevant additional information collected previously.

In some products, the data collected and recognition of the objects detected are used to predict the behavior of other ships. This prediction is used to avoid collisions and evaluate the COLREGs.

[1] In this study, camera systems which use the visible color spectrum are referred to as RGB camera systems (RGB stands for red, green and blue).

For route planning, the position and speed of the ship plus other ships around it and weather data are used to optimize route planning in terms of time or fuel consumption.

As part of situation awareness, data is obtained about the ship in order to examine this in more detail. To this end, other marine traffic or local weather conditions are recognized using a wide range of sensors, then contextualized and merged for an overall understanding.

What is important in terms of these use cases is a harmonized description of the resulting data sets for individual use cases, such as a *Maritime Perceived Environment* data set resulting from object detection. AI-based systems and also system verification can only be split into modules if there is this harmonization (Burmeister et al., 2020). Standard harmonization would also simplify the integration of multiple AI-based systems and verifying such integrated systems.

4.2. Summary of the Market Analysis and Consequent Need for Action

Looking at the state of development, there are clear differences in the scope, specification and use of AI modules in product functionality. In terms of the study, the market analysis helps identify basic information requirements and make appropriate recommendations. However, product descriptions alone allow merely an initial assessment of the extent to which the products being considered make use of AI technology.

Judging by the annual increase in patent applications over the last two decades, it can be inferred that the development of a Verification Guideline is currently very important and constitutes a prospective step towards market readiness for many manufacturers. In particular, the variety of AI-based systems identified as part of the market analysis suggests a large number of different technologies need to be examined.

In order to be able to cover the large number of different systems, it is worth designing verification procedures in a model-agnostic manner. It therefore follows that a generic view of AI-based systems and their applications and data sources is very important. Furthermore, it is impossible to predict which future AI-based systems will be used in MASS. These still unknown technologies could also be covered by a model-agnostic approach. To this end, a verification procedure which takes an evidence-based approach is proposed in the following sections, which focuses on processing input data and the resulting output data for the AI-based system. This is primarily intended to answer the question of “whether” the AI-based system works and not “how”.

4.3. Fictious AI-based System as an Example Use Case

A fictious example use case is introduced below, used to explain the proposed verification and Safety Guideline. The example is an image-based bearing sensor, whose specifications can be found in the product data sheet listed below. The explanations based on the example are framed in turquoise boxes and contain supporting explanations of how the results of the study can be transferred to verifying AI-based systems.

Product description:
The image-based bearing sensor evaluates a camera image with a focus on identifying and determining the orientation of ships in the area. The bearing is displayed in graphic form on a screen above the camera data relative to the center of the ship and provides an indication of this ship's orientation after evaluating the camera data.

Performance characteristics for the full system:
Recognizing 95% of all ships when all operational requirements are met. Maximum level of faulty detection of non-ship objects at less than 5%.

Camera features:

- Color profile: RGB
- Field of vision: 122°
- Image orientation to the center of the ship: 000°
- Focal length: 10 mm
- Image resolution: 2592 px × 1944 px
- Frame rate: 15 Hz

Operating requirements:

- Brightness: Daylight
- Maximum roll angle: 5°
- Maximum pitch angle: 3°
- Weather conditions: clear visibility (no precipitation, fog, spray, dust)
- Identifiable types of ship: Cargo ships

Installation:

- 1 screen on the bridge.
- 1 camera on the foreship mast with nothing visibly obscuring the lens.

Input:

- Video feed from the installed camera.

Issue date:

- Camera feed with overlay.
- Ship's bearing also displayed as a number in the overlay above the ships.
- Proprietary data stream in accordance with standard NMEA 0183 (DIN, 2011), similar to Radar Target Message.

5. Integrating the Verification and Safety Guideline

The study proposes a Verification Guideline (see section 6) for the BSH and a Safety Guideline (see section 7) for manufacturers. Both guidelines complement each other and cannot be considered one without the other. This integration ensures that procedures are already established in the Safety Guideline which prepare an AI-based system for the best possible outcome when it comes to the Verification Guideline. This means development and verification of the AI-based system can be pursued in a targeted, efficient manner.

The Safety Guideline is aimed at manufacturers of AI-based systems and has the following two objectives:

- Ensure the verifiability of the system.
- Development of a sufficiently safe system with the prospect of passing the audit.

It is required that manufacturers will not in fact only consider the Safety Guideline just before the audit, but instead right from early development of the AI-based system. This allows the foundations for a sufficiently safe AI-based system, capable of being audited, to be laid at an early stage.

The Verification Guideline is aimed at the BSH and the following objectives are pursued with regard to the AI-based system under consideration:

- Inspect that the information and safety technological functionalities properly.
- Reliable certification of AI-based systems.

The partnership between manufacturer and audit authority in the verification process is primarily expressed in the preparatory communications from manufacturer to audit authority. Preparatory communication before audit is what links the Safety and Verification Guidelines. This link is illustrated in simplified form in Figure 10.

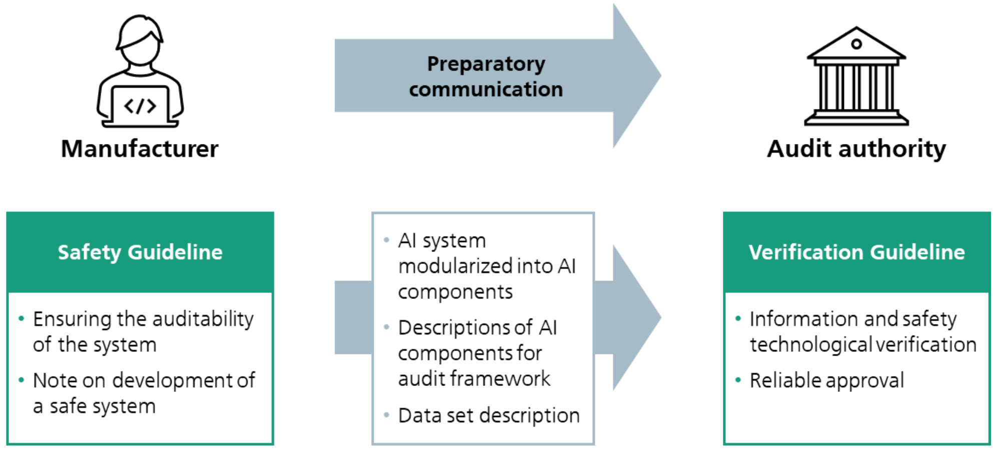

Manufacturer
Preparatory communication
Audit authority
Safety Guideline
• Ensuring the auditability of the system
• Note on development of a safe system
• AI system modularized into AI components
• Descriptions of AI components for audit framework
• Data set description
Verification Guideline
• Information and safety technological verification
• Reliable approval

Figure 10: Preparatory communication from manufacturer to audit authority.

For manufacturers, the content of this preparatory communication before audit derives from the Safety Guideline. Preparatory communication essentially comprises the following elements:

- The AI-based system, which is being audited, split into AI modules[1] (see section 7.1 subsection F1).
- A description of the AI modules with specifications to define the audit framework (see section 7.1 subsection F2).
- A data specification for obtaining suitable test data (see section 7.3 subsection D2).

Pre-audit preparatory communication is vitally important for conducting any verification and obtaining certification from the BSH.

This study considers only those AI-based systems where the AI model is frozen. The reason for this is that the performance of non-frozen models may change after verification and certification and so verification and certification would prove meaningless.

The following is a basic introduction to the use of the Safety and Verification Guideline and the communication between manufacturer and audit authority. More detailed explanations can be found at the end of the sections on the Safety Guideline (section 7) and Verification Guideline (section 6).

The manufacturer plans to develop a product that includes an AI-based system. The product is an image-based bearing sensor. By taking the Safety Guideline into account, the manufacturer is helped in two respects.

Using the Safety Guideline to develop a safe AI-based system

Firstly, recommendations are made which support development of the AI-based system. The recommendations represent proven development practice and make an AI-based system as safe as possible. The Safety Guideline focuses on data quality. In the example of the image-based bearing sensor, mostly RGB image data and AIS data are used. Accordingly, the data quality information refers to these data sets. Following this recommendation improves the chances of the AI-based system passing the verification and being certified.

Communication with the audit authority

Secondly, actions necessary to prepare for the audit are explained, which the manufacturer must guarantee. Taking these actions ensures that the manufacturer’s AI-based system can be verified. The first action is splitting the AI-based system into modules. The manufacturer would break it down into individual AI modules which can be verfified. One AI module in the example use case might be ship recognition in image data. As a second action, the manufacturer would define all AI modules in terms of their scope of application, i.e. under which conditions the ship recognition module needs to work. The final action would be for the manufacturer to define those data sets on which the AI modules were trained. In the example use case, the RGB image data set and the AIS data set would need to be defined.

[1] AI modules are sub-units of the AI-based system to be tested which can be tested individually (see section 7.1 subsection F1).

Using the Verification Guideline to reliably verify the AI-based system

An audit procedure is proposed to the audit authority, which permits a reliable safety and information technological verification. In order to carry out the image-based bearing sensor verification, the manufacturer must follow the preparatory communication.

Firstly, this involves the AI-based system being submitted to the audit authority broken down into modular form. The auditor is then able to audit the AI modules, such as ship recognition, individually.

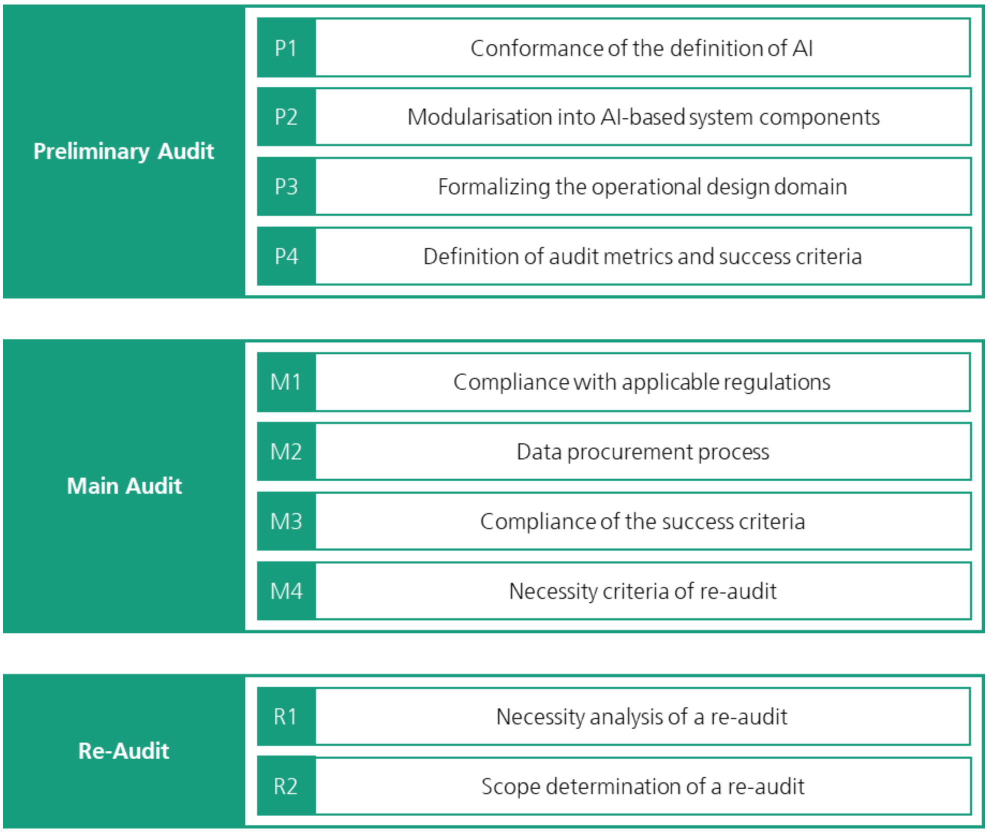
Secondly, the manufacturer submits the specifications for the AI modules to the audit authority, from which the verification framework is derived. The verification framework includes the operational domain for the AI modules and measurable criteria for evaluating a successful verification.

Thirdly, a description of the data sets used is submitted to the audit authority. This should show the audit authority what data can be used to verify the AI modules.

6. Verification Guideline

The objective of this Verification Guideline is to use a sequence of audit steps to verify AI-based systems in a model-agnostic manner in terms of whether the data and safety features work properly, and to be able to approve them with confidence. The focus lies in establishing "whether" and not "how" an AI-based system works.

The sequential structure of the Verification Guideline is shown in Figure 11 and shows the audit broken down into three sections (Preliminary Audit, Main Audit and Re-Audit) and the audit steps. One audit step includes tasks to be performed by the audit authority as part of the Verification Guideline. The sequential structure was chosen as part of the study in order to separate out individual parts of the Verification Guideline from a technical point of view and to provide the audit authority with guidelines which are easy to follow and which inherently dictate the structure of the verification process.



Preliminary Audit	P1	Conformance of the definition of AI
	P2	Modularisation into AI-based system components
	P3	Formalizing the operational design domain
	P4	Definition of audit metrics and success criteria
Main Audit	M1	Compliance with applicable regulations
	M2	Data procurement process
	M3	Compliance of the success criteria
	M4	Necessity criteria of re-audit
Re-Audit	R1	Necessity analysis of a re-audit
	R2	Scope determination of a re-audit

Figure 11: Verification Guideline with sections and steps.

The audit sections (Preliminary Audit, Main Audit and Re-Audit) form a theme-based structure in order to bundle together certain aspects of verifying AI-based systems and combine them into specific audit steps. An audit step is the smallest unit of audit and represents a specific technical task that must be carried out as part of verification and certification of an AI-based system. Audit steps are carried out in sequence and each one must be passed in terms of the task and its purpose. If any audit step cannot be fully carried out within an audit step due to the AI-based system and any results of the test, the test is considered not passed and rework by the manufacturer is necessary.

6.1. Preliminary Audit

The preliminary audit includes audit steps to assess the need for this specific test of AI-based systems. By using explicit differentiation, only systems that fall within the definition of an AI-based system should be examined. Furthermore, the basic ability of the AI-based system to be verified is established as part of the Preliminary Audit. Establishing auditability is necessary so that the model-agnostic verification procedure can be applied to the AI-based system.

P1 | Conformance of the definition of AI

Verifying the AI parts of an AI-based system separately was proposed by introducing test module K (see Figure 8) into the verification and certification framework. To be able to use existing processes effectively, the verification of an AI-based system must begin by checking whether the system meets the definition of an AI-based system within the scope of the study. Section 2.1 provides an overview of the spectrum of these technologies and corresponding definitions. AI-based systems that make use of the technology listed therefore meet the definition of AI and are considered in the processes set out below. It should be noted that the definition of an AI-based system may change over time due to ongoing developments.

Based on the described functionality, the scope of features in the example use case of the image-based bearing sensor suggests a system that can be described as an AI-based system in terms of this study. In view of current developments, it can be assumed that recognition is based on a CI approach (see section 2.1) in order to recognize shapes in general that correspond to a ship. Similar approaches use models that have been trained to recognize these using a large quantity of images of ships. The system's knowledge therefore does not consist in the symbolic description of ships, but in mathematical image evaluation in order to recognize known patterns and be able to assign these to known classes of objects.

P2 | Modularization into AI-based system components

Considering the variety of AI-based system architectures and the potential complexity of these systems, holistic verifying is only possible in certain cases. It is also not practical to verify AI-based systems using procedures specific to the architecture. Modularizing the AI-based systems into specific components that solve specific tasks can be what makes verification possible in the first place. This reduces complexity to the extent that specifying the functionality and procedures relevant to the audit can be carried out one component at a time.

In the context of the study, a system is considered modularized if the corresponding number of outputs can be unambiguously ascribed to each set of inputs. Matching input to output values forms the basis for evidence-based verifying of AI-based systems. Figure 12 shows schematically modularization can be achieved by unambiguously matching input and output data. Values E1 to E7 and A1 to A7 represent input and output values and K1 to K3 represent the system modules. Modularization therefore happens in two stages. In the first stage, corresponding input and output data streams for the AI-based system are split up and in the second stage the AI-based system is modularized into AI components based on the classification of the data stream.

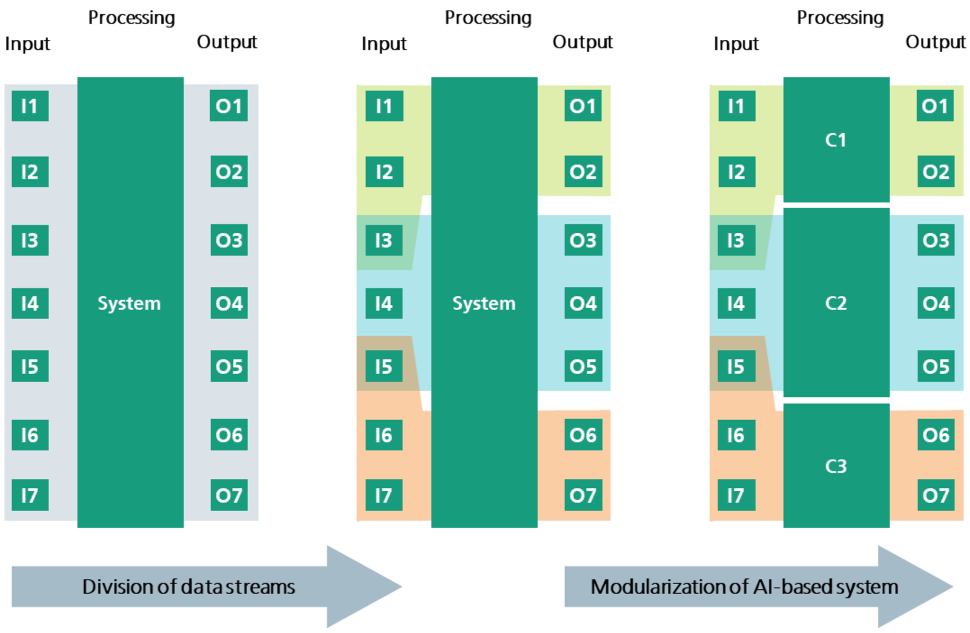


Figure 12: Breaking down AI-based systems into modules. I stands for input, O for output and C for system component.

The full verification procedure uses the input, processing and output (IPO) principle to carry out evidence-based verifying. Evidence-based here means that the system is examined based on observation and evaluation of its (reproducible) performance. Only the input and output data are known to the audit authority. In terms of the study, the processing unit (AI component) is treated as a black box from a model-agnostic perspective and no more detail is required. Treating it as a black box is both in the interest of manufacturers to keep new technology under wraps and in the interest of the audit authority to treat the large quantity of current and potential architectures not individually, but as a whole, i.e. model-agnostic. Treating it as a white box would complicate verification and certification and would only be possible for sAI models anyway. With CI models, processing is not comprehensible. By utilizing the IPO principle in the verification of AI-based systems, this remains practical and scalable. In the following, the unit to be verified will be referred to as an AI module.

Separating the image-based bearing sensor out into a module depends on how far individual mathematical processes have been integrated (recognition and calculating the bearing).

To illustrate breaking down into modules, it is assumed that ship recognition and bearing estimation are two separate modules and can be verified individually. The manufacturer provides both modules with their input and output interfaces in an executable environment.

The following diagram (Figure 13) represents the AI-based system broken down into modules based on Figure 12.

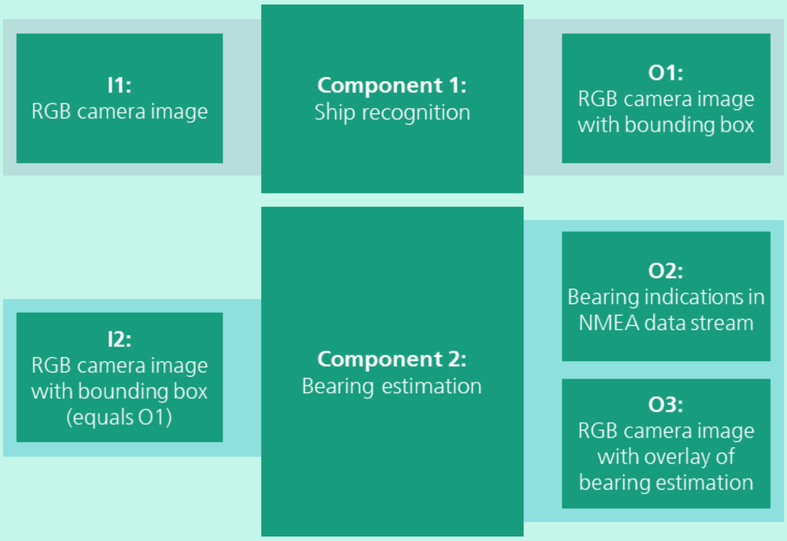

I1:
RGB camera image
Component 1:
Ship recognition
O1:
RGB camera image with bounding box
O2:
Bearing indications in NMEA data stream
I2:
RGB camera image with bounding box (equals O1)
Component 2:
Bearing estimation
O3:
RGB camera image with overlay of bearing estimation

Figure 13: Possible breakdown into modules for example use case.

The ship recognition module reflects the technology responsible for identifying ships in a continuous data stream of images at a resolution of 2592 px × 1944 px and a frame rate of 15 Hz. Output is the RGB camera image, as well as a tagged list of bounding boxes that were recognized by the module within the image.

The bearing estimation module receives the camera images and a list of bounding boxes, which are converted into a list of bearings with orientation information in degrees and also outputs a composition showing the images and bearing information by means of a graphical interface. The module also generates a NMEA-compliant data stream to make the sensor results accessible to other systems.

It is essential for the audit authority to ensure that it is possible to use the modules for the audit steps which follow, and to evaluate results by entering data in accordance with the data description.

P3 | Formalizing the operational design domain

In order to be able to delineate the functionality of an AI component and to set the framework conditions for the main audit, the manufacturer must clearly specify the framework conditions under which the AI component should function. This delineation, defined as an operational design domain, specifies the variables to which the functioning of the AI module can be restricted. The spectrum of values within which the module receives inputs as defined by the manufacturer and generates corresponding outputs is defined as the operating envelope. Accordingly, the module can be used within this range of values. It cannot function outside this range of values and should not produce any output in order to avoid unexpected decisions or actions.

Each AI module must fall within a defined operational design domain, which both reflects the capacities of the AI module and creates sensible boundaries within the scope of application for the AI module. So the operational design domain provides a basis for the manufacturer's data description, which includes the variables needed for the main audit in order to be able to obtain data for verifying the AI module.

The operational design domain can be formalized using recognized methodologies and should align with generally accepted forms. The operational design domain tool originally came from the automotive industry and is used to describe potential areas of application for (semi-) autonomous vehicles (Gyllenhammar et al., 2020). The Operational Envelope concept is proposed for the maritime field, prospectively for MASS. This is based on the Operational Design Domain. It also takes into account the responsibilities and interfaces both within and between the realms of autonomization and human operators (Rødseth et al., 2022). Regardless of the method of formalization chosen, the operational design domain should meet the following criteria to ensure an AI module can be verified.

- For each AI component of an AI-based system the operational design domain must be defined and provided.
- The operational design domain clearly defines and demarcates all input values used within the AI component and defines fixed value ranges.
- For each input value, it must be clearly defined whether and to what extent it is used by the AI component, and which input values were assumed when developing the component. Similarly, it must be clearly defined which values are not allowed to be processed and how the component handles these.
- The outputs from the AI component should be defined in a similar way to the input values. An approach based on the *Maritime Perceived Environment* (Burmeister et al., 2020) might be useful here.

When designing a framework for formalizing and verifying operational design domains, it is worth setting up a central database for operational design domains that have already been verified. An open-access method for defining and generalizing operational design domains introduced by the Association for Standardization of Automation and Measuring Systems (ASAM, 2021) and based on the Operational Design Domain can also be adapted for verifying AI-based systems in the maritime sector. Grouping together operational design domains that have already been verified enables the audit authority to identify common areas and take these into account as a basis for future audit. Open access to such a database can be used by manufacturers of AI-based systems as a standard when formalizing the operational design domain.

By splitting the image-based bearing sensor into two components, two operational design domains must also be defined.

The operational design domain for the ship recognition component is mainly characterized by the technical operating requirements and is already largely defined by these.

The operational design domain for the bearing component is based around a dependency on the ship recognition component and its output data. So the operational design domain needs to take into account any domain boundaries for the output data provided by the ship recognition component.

P4 | Definition of audit metrics and success criteria

Once the operational design domain has passed this test, the audit authority must then evaluate the audit metrics defined for the AI module and verify these in terms of the intended function of the AI component. In this context, audit metrics means directly measurable or indirectly determinable variables that are used to evaluate the output of an AI component and make the results quantifiable and comparable.

Audit metrics must enable the output values from an AI component to be evaluated in terms of quality of the output values. The intended functionality of the AI component and so the form taken by the corresponding output values must be taken into account. In the simplest scenario, output values can be quantitative values (e.g., non-discrete numerical ranges, target category values or two-value logic statements) or those with qualitative characteristics (e.g., recommended navigation action). Quantitative audit metrics can, for example, be determined using various methods such as a confusion matrix or measuring Euclidean distance to find the similarity between two numbers.

When defining the audit metrics, the manufacturer must also define success criteria for the function of the AI component during the development process. Defining the success criteria is related to the operational design domain and gives the audit performance boundaries that must be adhered to by the AI component.

The following requirements should be met when defining success criteria:

- The audit metrics give the audit authority the opportunity to evaluate, by looking at input and output values, whether the AI components works (as well) as described by the manufacturer.
- Where standards or norms for audit metrics and success criteria already exist for similar AI components, then the success criteria specified by the manufacturer may be equally or more critical.

It is recommended that the audit authority, in conjunction with other test institutions, establish international standards or norms using appropriate audit metrics and success criteria for AI components. This can mean less effort for manufacturers and regulatory bodies alike. The manufacturer might use existing standards or norms and the audit authority might introduce scalable verification procedures.

In the example use case, the audit metrics and associated success criteria are divided between both AI components.

The ship recognition component should be assessable using audit metrics that reflects the proportion of ships classified correctly and incorrectly. This can be achieved, for example, by using a confusion matrix, as shown in Table 4, and deriving variables from that (Navlani et al., 2021).

Table 4: Example of a confusion matrix as the basis for audit metrics in the example use case.

	Ship identified: YES	**Ship identified: NO**
Ship nearby: YES	**True positives: 980**	**False negatives: 4**
Ship nearby: NO	**False positives: 1**	**True negatives: 265**

The audit metrics for the bearing component are defined by the accuracy of the estimated bearing from the image analysis and a reference from AIS or radar data. The required accuracy should be checked against the manufacturer's specifications or determined some other way. In the example use case, if the manufacturer has not specified any criteria as to how high they feel the accuracy should be, a sensible value should be assumed. In this case, it makes sense to define the accuracy by checking against another device for determining bearing and returning the bearing with a high degree of agreement.

6.2. Main Audit

All audit stages carried out as part of the main audit focus on the actual examination of the AI component of the AI-based system and determining whether it works properly. This audit section includes investigation of legal compliance, creating a test data basis for the verification, and verifying and evaluating according to defined audit metrics and success criteria. Finally, conditions for a renewed audit ("re-audit") are defined.

M1 | Compliance with applicable regulations

Due to the wide range of guidelines, standards and legal frameworks currently being developed that deal with the regulation of AI-based systems, it is difficult to define a fixed set of rules on which a Verification Guideline can be based. A review of the detailed regulations (see section 3.5) shows how extensive and dynamic the draft documents are and how differently they can affect the certification and operation of AI-based systems.

An examination of existing regulations, i.e., documents that define AI (sub-)systems, specify performance and standardize operating variables, can only be established in theoretical terms as part of a Verification Guideline. Standards that relate to the architecture and structure of AI-based systems and which fall within the definition chosen in this study could provide contradictory statements and recommendations for the certification and development of these systems. This might particularly be the case with AI technologies that are already very advanced and have already been experimentally approved in other sectors of industry. This includes image recognition systems trained using a CNN (DIN, 2020).

The audit authority should therefore be able to include existing requirements in these procedures and, if necessary, specify recommendations that vary from the audit steps listed here. Actual separation requires a decision on a case-by-case basis and should therefore be individually coordinated for each AI component that is affected by it.

From the audit authority's perspective, it is crucial to constantly review new regulations and adapt the verification procedure to these changes. Due to rapid pace of change in the field, this process should be carried out regularly.

With the example use case, review of existing regulations applies mainly to the output data in a format according to NMEA 0183 (DIN, 2011) as well as the depiction of the graphic elements on the bridge (IHO, 2014).

M2 | Data procurement process

At the core of evidence-based verification of AI components lies the comparison of input and output data against the operational design domain. By entering data that the component does not recognize, the audit authority should be able to prove whether the AI component is functioning properly.

Using a description of the input and output data from the manufacturer, the audit authority must be able to enter data into the component in a form the component expects. This description of the data, introduced as a definition of data as part of the study, serves as a mechanism for formalizing the interaction between manufacturer and audit authority. The aim is then to obtain the data independently. Possible ways for the manufacturer to set up a data description are listed in section 7.3 subsection D2.

When the manufacturer submits a data description, it must first be checked for completeness as part of the certification process. In the first step, the audit authority must ensure that the same variables are used in the data description and in the formalized operational design domain. This ensures that, with reference to the formalized operational design domain, it can be determined from which data space new data must be obtained. If the audit authority determines at this stage that the data description is incomplete or does not correspond to the variables from the operational design domain, the verification procedure must be halted and the manufacturer must amend the data description. Two possible scenarios were identified for the data acquisition process, which require different actions on the part of the audit authority:

- Data meeting the data description already exists: The first scenario assumes that data for testing already exists and matches the manufacturer's data description. This applies particularly where a component with the same or similar functionality has already been approved and the relevant data has been obtained by the audit authority. This requires a database with valid input data and matching operational design domains. It makes it easier to compare components with the same function but different standard application. In principle, storing data depending on the application, such as by using operating envelopes, is recommended. Random, unstructured creation of so-called "data dumps" is not to be encouraged.
- Data meeting the data description does not already exist: This scenario requires obtaining data using data generation techniques, or using data which is publicly or commercially available. Procedures for obtaining data can be managed in-house by the audit authority or transferred to external service providers. When obtaining the data, it must be ensured that it was not also used by the manufacturer when developing the AI component.

Data generation is considered a promising process for data acquisition. Advantages are as follows: Firstly, data is obtained that was not used by the manufacturer in development, secondly, the possible amount and variety of data that can theoretically be generated exceeds that of the data available elsewhere and, thirdly, generated data can in fact be stored with target values (Nikolenko, 2021a). When generating data, a fundamental distinction should be made between two paradigms: Data augmentation and data synthesis. With data augmentation, new data is generated based on existing data (Nikolenko, 2021b). This is achieved with image data, for instance, through transformation processes (symmetrical operations such as reflections or rotations) or through enrichment processes (additions to objects). With data synthesis, however, totally new, i.e., artificial, data is generated. Both data augmentation and data synthesis are in fact widely used. Data augmentation is used, for instance, with image data that is supplemented by artificial objects (Ekbatani et al., 2017; Frid-Adar et al., 2018). Data synthesis involves synthesizing image and video data from game engines (Korakakis et al., 2018; Tsirikoglou et al., 2017) or other forms of data such as time sequences (Zhang et al., 2018). Data synthesis can also be implemented using declaration languages that can simulate any scenario. This means numerous scenarios with specified variations can be generated at any time. The advantage of this approach would be to avoid data dumps, because only the command sets for generating scenarios would have to be saved, and not the resulting data set. The development and implementation of this approach is currently the subject of research at Fraunhofer CML.

With training models, it turns out that using a dataset expanded through data augmentation improves model performance more than a purely synthetic dataset, because an augmented dataset has greater variability (Seib et al., 2020). It is assumed that testing with an augmented data set is more effective than with a purely synthetic or exclusively real-world data set. Published standards (see section 4.1.1) for the targeted data sources can be helpful in data generation, as they can specify possible formats and values for the data to be generated.

The audit authority is responsible making sure only data not used by the manufacturer is used for this testing. Where the manufacturer has indicated that publicly or commercially available data sets were used for training, the audit authority must ensure that it does not use the same. The audit authority can ensure this by using data synthesis or augmentation techniques, or more generally by not using open-access data sets.

Once the data has been obtained, it must be checked in consultation with the manufacturer to see if it corresponds to the manufacturer's expectations when compared with the data description submitted. The process shown in Figure 14 provides a summary of how the coordination between audit authority and manufacturer can be agreed when it comes to the data description.

The process proposed allows for interaction between audit authority and manufacturer regarding two points. On the one hand, the manufacturer can improve the completeness of the data description if the audit authority has identified shortfalls compared to the operational design domain. A second check is performed after obtaining sample data that corresponds to the manufacturer's data description. This step gives the manufacturer the opportunity to compare data generated by the audit authority with in-house expectations and advise possible variations in order to make adjustments to the data description if appropriate. Variations in the test data obtained by the audit authority or manufacturer can, for example, be due to statistical distribution, noise or more generally false assumptions (e.g., cognitive bias).

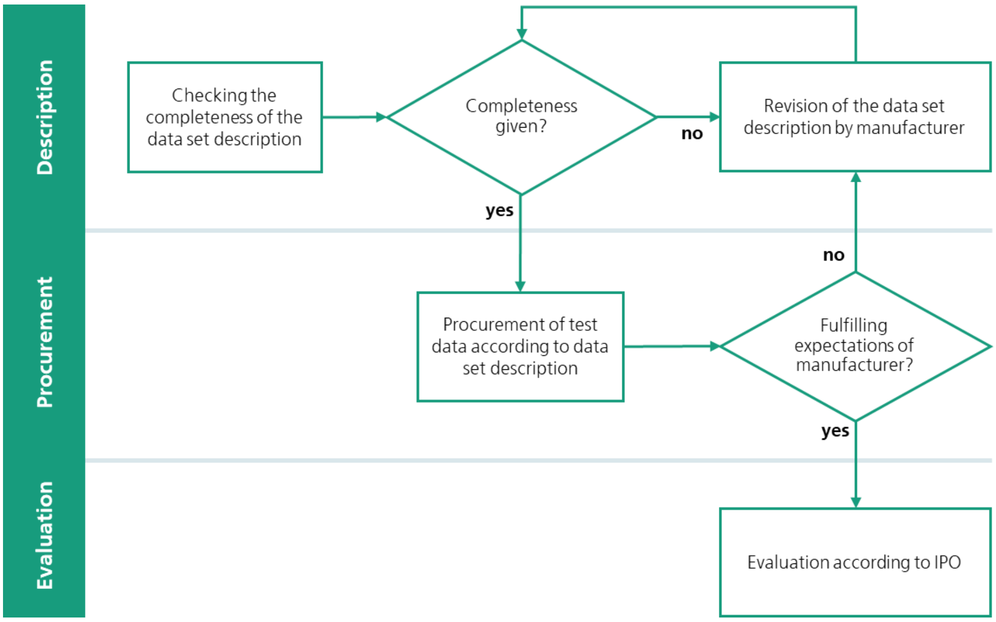

Description
Procurement
Evaluation
Checking the completeness of the data set description
Completeness given?
no
Revision of the data set description by manufacturer
yes
no
Procurement of test data according to data set description
Fulfilling expectations of manufacturer?
yes
Evaluation according to IPO

Figure 14: Iterative coordination for data procurement between audit authority and manufacturer.

Communication between manufacturer and audit authority when obtaining test data based on a data description offers the possibility of detecting errors in the performance of the AI component at an early stage. The manufacturer's detailed description of the expected input and output data increases the expectation that any test data set has been thoroughly explored in the development process. Furthermore, a strictly formalized data description between manufacturer and audit authority has the advantage that ambiguities or misunderstandings in the description of the data or functionality of the component are kept to a minimum, as ambiguities and scope for interpretation are avoided.

When it comes to the image-based bearing sensor, this data acquisition process focuses on acquiring image data which corresponds to the characteristics of the ship recognition component and reflects real-world images of ship-to-ship situations.

Since such systems may be being verified for the first time, no data is available at the time of verification and must be generated by the audit authority. This process can be achieved either by proficient companies specializing in generating image data digitally, or by building in-house expertise.

Current research results in the field of AI-assisted image generation can continue to develop specialized support systems that enable the necessary data to be generated. Systems such as the "DALL-E 2" AI model are currently demonstrating great success in the programmed generation of realistic image data, which could also fundamentally be used to generate data products to approve the example use case. Examples of artificial image sequences (Figure 15, Figure 16 and Figure 17) that were generated using DALL-E 2 are listed below (Ramesh et al., 2022).

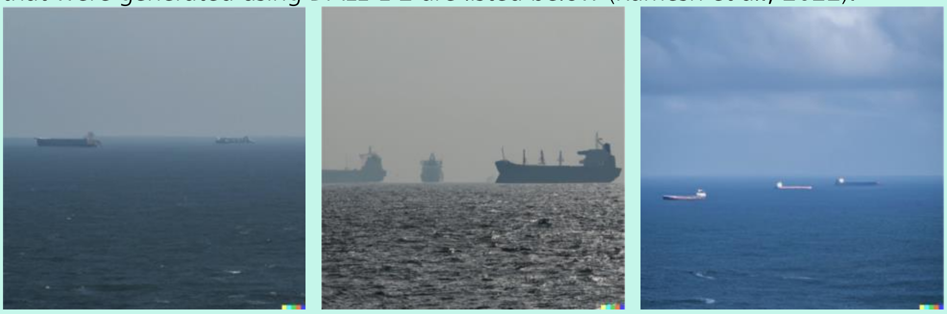


Figure 15: Artificial images for the prompt "fleet of ships on the horizon".

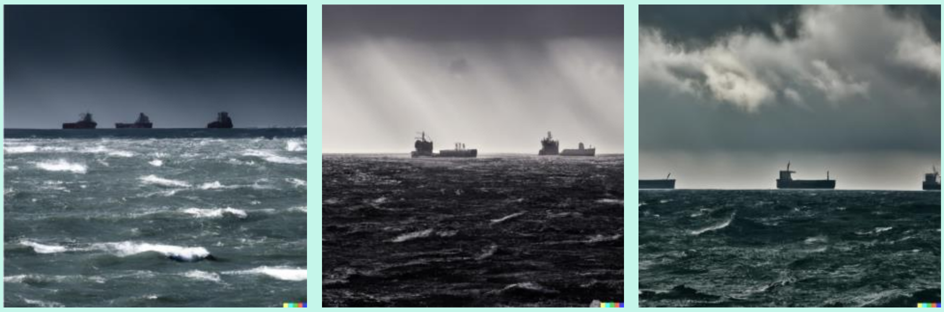


Figure 16: Artificial images for the prompt "fleet of ships on the horizon during a storm".

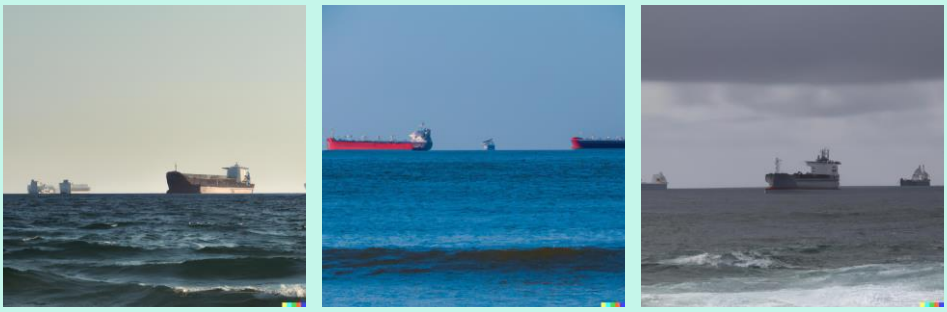


Figure 17: Artificial images for the prompt “ships on the horizon looking towards the camera”.

Obtaining the data necessary for the bearing component fundamentally depends on the data used for the ship recognition component. When obtaining data for testing the ship recognition component, care must be taken to take the operational design domain of the bearing component into account. It should basically be possible to test the systems separately. A combined test might be seen as one way of accelerating the process and improving the efficiency of the Verification Guideline.

M3 | Compliance of the success criteria

The principal aim of the audit procedure is to validate the success criteria that were formalized by the manufacturer as part of the preliminary audit. This validation is applied along the entire operational design domain. A check should be made outside the operational design domain, to see whether the AI component is inactive as expected. This step reflects the actual information and safety verification for the AI component.

Data generated as part of the data acquisition process can be entered into the components in the executable environment and the outputs recorded. Meeting the success criteria depends on the degree to which the test results match the manufacturer’s success criteria.

In the example use case, the number of correctly identified ships must be in the range of 95% of the test data. Furthermore, no more than 5% of non-ships may be incorrectly identified as ships. This must be guaranteed for all test data generated regarding the operational design domain.

M4 | Necessity criteria of re-audit

If an AI component passes the audit, the audit authority must determine the conditions for a re-audit (see subsection N1). The criteria qualifying for a re-audit can be time- or event-based and apply to individual components. So, the need for a re-audit can be determined one component at a time.

Examples of time-based re-audits include legally mandated periods during which the function of the AI component must be re-audited. The reason here is that in the Verification Guideline, frozen AI-based systems, more particularly their models, and the operational design domain are viewed as unchanging. Over time, this can lead to drift between the model and reality (see section 2.1.1). In order to be sure that the components are staying up to date, it is essential to test regularly for proper functioning. For example, the EU AI Act mentioned in section 3.5 provides for periodic conformity assessment for approved AI-based systems.

Event-based re-auditing, on the other hand, is predicated on direct changes to the AI component and possible impact on operating safety. In particular, following software updates or changes to the hardware, the audit authority must determine whether the function of the AI component validated during the test has been affected.

The re-audit criteria are determined at the end of the verification process and are determined regarding the operational design domain. The function of the example use case fundamentally depends on the characteristics of the RGB camera. Any change to this hardware component would lead to a re-audit.

The bearing component inherently depends on the characteristics of the ship recognition component as well as the installation data used to estimate the bearing. Design changes to the camera must also be considered when it comes to the bearing components. These may be listed as a performance feature in the product data sheet, but nevertheless lead to a re-audit.

If the ship recognition component is expanded to detect objects that were not included in the original scope (such as sea marks or obstacles), the frozen model can no longer be used and must be replaced with a new model. The new model must be tested again to demonstrate that the functionality claimed for the new classifications as well as the full previous spectrum of performance continue to be valid (see section 2.1.1).

6.3. Re-Audit

The need for a re-audit may arise from internal changes to the AI-based system, external changes in the operational design domain, or the timing of certification. The needs assessment and scope determination are listed in the following subsections.

R1 | Necessity analysis of a re-audit

The audit authority must keep a list of all approved AI components to determine the necessity for a re-audit. This is particularly necessary where the component can no longer be guaranteed to operate reliably due to the occurrence of some external influence (see section 6.2 subsection M4). Changes to the software need to be reviewed thoroughly to see if the AI component needs to be tested again or if these can be validated using simplified procedures before being installed on the component.

R2 | Scope determination of a re-audit

Once the need for a re-audit has been determined, the audit authority also needs to determine the scope. Depending on the scope, the conditions for a re-audit specified during certification mean that a single AI component, several (connected) AI components or the entire AI-based system may have to go through the verification procedure again.

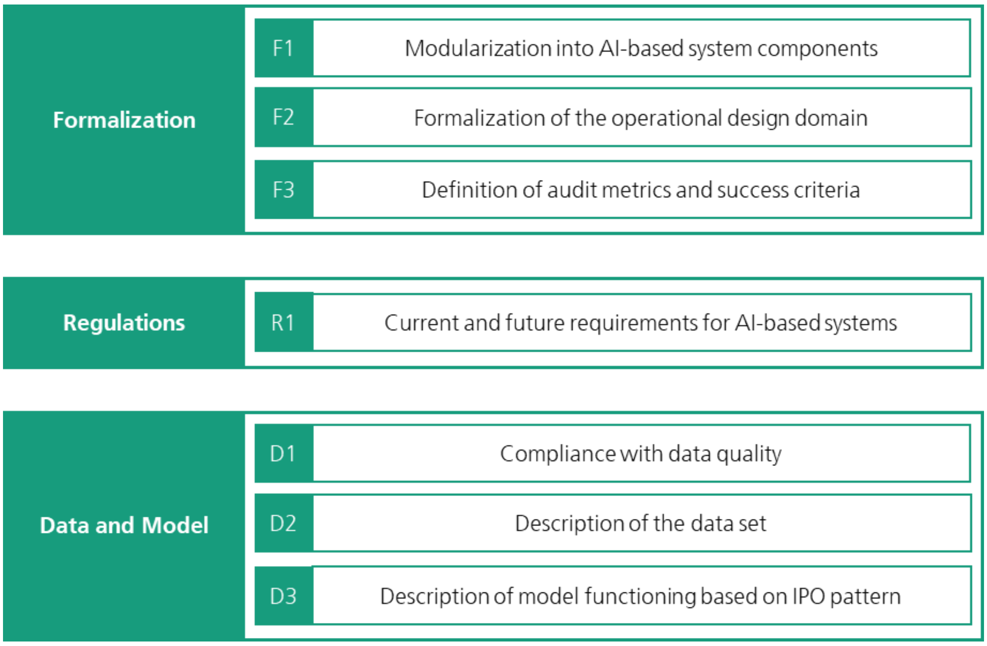
A re-audit is considered as passed once the required system outputs are met, the same as for the success criteria set out in section 6.1. Whenever a re-audit is carried out or the functionality of an AI component changes, the criteria for a re-audit must also be reviewed and revised.

7. Safety Guideline

The Safety Guideline is intended for manufacturers of the AI-based systems to be audited and serves two aims:

- Making sure the system is auditable.
- Providing information used for development of a sufficiently safe system with the prospect of passing the audit.

In order to achieve these two aims, the Safety Guideline was developed in alignment with the Verification Guideline. The Safety Guideline consists of three sections, each with individual steps, although the steps do not necessarily have to be followed in the specified order. The sections and steps are set out below in Figure 18.



Formalization
F1 Modularization into AI-based system components
F2 Formalization of the operational design domain
F3 Definition of audit metrics and success criteria

Regulations
R1 Current and future requirements for AI-based systems

Data and Model
D1 Compliance with data quality
D2 Description of the data set
D3 Description of model functioning based on IPO pattern

Figure 18: Safety Guideline with sections and steps.

The first section, formalization, is intended to get an AI-based system ready for audit. This includes steps whereby an AI-based system can be adequately described so that it can be tested by the audit authority. To this end, the entire system is described one component-by-component in terms of its operational design domain and expected performance.

In the regulatory section, the manufacturer is shown how to deal with both the current and prospective regulatory framework.

In the final section, Data and Model, steps are explained relating not only to data quality but also to the description and reproducibility of system performance. For the first of these, instructions are given for proper, qualitative processing of data during the development process. For the other, common mathematical and technical methods are listed, which make it possible to describe the data set used in development in a manner easy to understand, as well as describe the performance of the AI model developed.

7.1. Formalization

F1 | Modularization into AI-based system components

The manufacturer's AI-based system is verified in modular form. For the manufacturer, this means that an AI-based system must be adequately split into individual AI components during preparations for the audit (see section 6.1). The manufacturer can lay the foundations for this at an early stage by choosing a system architecture suitable for verifying where the AI-based system is broken down into components.

For the manufacturer, verifying in modular form comes with the advantage that test results for the system can be obtained one component-by-component. If improvements to the system are required, these can be carried out specifically on the components requiring improvement. The manufacturer can also carry out improvements on an individual component independently, which the audit authority then also verifies individually.

Ultimately, the unit under test will be referred to as an AI component.

The image-based bearing sensor is an AI-based system that can be broken down into individual AI components by various functions. The audit authority would then verify the AI components individually.

The manufacturer might break the AI-based system down into two AI components as follows:

- Ship recognition component: Identifying ships using RGB image data.
- Bearing component: Estimating the ships bearings.

The manufacturer sends the AI-based system, modularized into AI components, to the audit authority as part of the preparatory communication for the audit (see section 5). One way of modularization is illustrated in Figure 13.

F2 | Formalization of the operational design domain

From the manufacturer's perspective, the operational design domain is the area of application in which the AI component is intended to function. The manufacturer is aware of the operational design domain for each AI component right from the development stage. Specifically, this is the input or output data space which the AI component can process or output as a result.

The formalized operational design domain is used by the audit authority to demarcate the audit framework for the AI component. The operational design domain needs to be formalized so that the audit authority can implement this as correctly and accurately as possible.

Techniques for formalizing the operational design domain are set out in section 6.1 subsection P3.

Both AI components as set out above must be formalized by the manufacturer in terms of the operational design domain. Using the example of the ship recognition component, the manufacturer must describe the image data used for development and used in the application from both a technical and content perspective.

The technical perspective refers to RGB camera system variables such as image resolution, focal length, field of view and similar properties. This is necessary so that when auditing AI components, the same image technology is used as in the application to reproduce image data that can be processed by the AI component.

The image content featuring in the application must also be described. This includes the circumstances, such as the weather conditions when the picture was taken. It can be seen from the product data sheet (see section 4.3) that the image-based bearing sensor only works in daylight in clear visibility.

From the formalization of the operational design domain, the manufacturer can see which images the image-based bearing sensor is given.

The manufacturer sends the operational design domain to the audit authority as part of the preparatory communication for the audit (see section 5).

F3 | Definition of audit metrics and success criteria

In a final step in the formalization process, the manufacturer must submit audit metrics and success criteria to the audit authority which can be used to assess AI component output quality. In this respect, the manufacturer must propose a method for measurement (audit metrics) as well as the standard to be attained (success criteria). These are reviewed in conjunction with the audit authority to see whether they meet the requirements of the audit authority. Where standards or norms for success criteria already exist for similar AI components, then the success criteria specified by the manufacturer may be equally or stricter. The audit metrics and success criteria must be defined individually for each AI component.

When developing an AI model, the manufacturer should use measurement techniques to measure and perfect the model’s performance. The choice of a suitable measurement technique depends primarily on the type of output. For instance, with binary or multi-class classification problems, using a confusion matrix is appropriate (Navlani et al., 2021).

The manufacturer should provide audit metrics and success criteria for each component so that the audit authority can evaluate the performance of the AI components and approve them if the results merit this.

The manufacturer can use a confusion matrix for the ship recognition component, as it is a binary classification problem (Navlani et al., 2021). If an image includes a sufficiently identifiable ship, then it must be successfully identified by the ship recognition component. The manufacturer must specify success criteria which must be met for the AI component to be considered functional in terms of safety and information technology. The manufacturer would normally formulate these parameters during development of this AI component for improvements to be measurable.

The manufacturer sends the audit metrics and success criteria to the audit authority as part of the preparatory communication for the test (see section 5).

7.2. Regulations

R1 | Current and future requirements for AI-based systems

During the development process, the manufacturer is responsible for reviewing and evaluating existing requirements for the selected architecture and operational design domain. Where AI technology has been selected to solve a problem, several requirements may exist that may have varying degrees of influence on the development of an AI-based system.

Due to lack of standardization among AI-based systems and underlying architectures, a detailed overview is only possible to a limited extent within the scope of this study. It is essential to consult with the audit authority during the development phase regarding which regulations currently apply to the verification procedure.

The manufacturer tests its AI-based system during development and preparations for the audit to see whether it must comply with any regulations and if so which.

Because the output of the bearing component falls under NMEA 0183, the manufacturer must comply with this (DIN, 2011). The manufacturer should also monitor progress on implementation of the EU AI Act (Europäische Kommission, 2021), as this could have an impact on its product (see section 3.5.2) and keep up to date with the guidelines for the development of deep learning image recognition systems (DIN, 2020).

7.3. Data and Model

D1 | Compliance with data quality

Poor data quality generally manifests as missing, incomplete, inconsistent, inaccurate, or duplicate data (DIN, 2020; Gudivada et al., 2017). Other types of poor data quality can often be observed in machine learning specifically: the use of too many variables, variables that are highly correlated with one another, or outliers in the data set.

Data quality is of major importance when it comes to machine learning. This is because usable data represents the knowledge on which the AI model is trained (see section 2.1.1). If data quality is poor, problems such as high bias or high variance can arise, which manifest themselves in the form of poor accuracy or precision. Figure 19 shows these effects in simplified form using throws of a dart. Poor data quality can mean that an AI-based system does not perform reliably, even if the machine learning methods chosen are highly suitable and the rest of the development follows high quality standards.

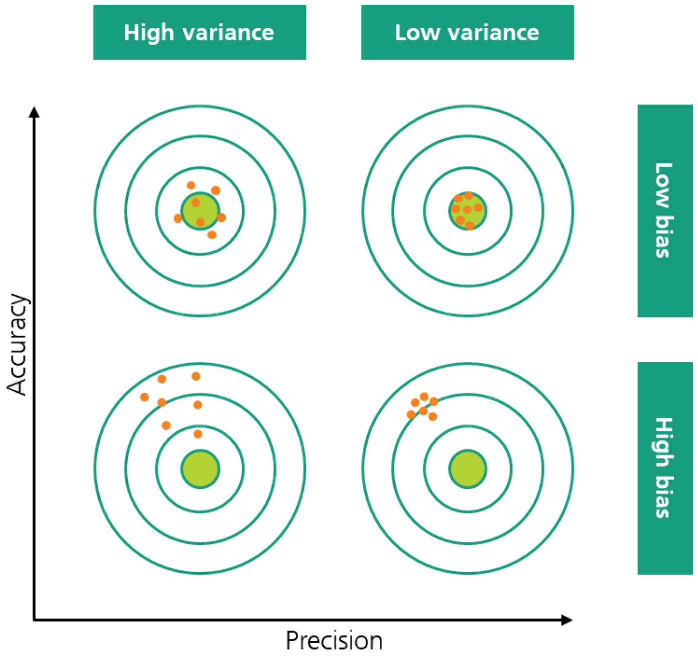


Figure 19: Relationship between variance and precision as well as bias and accuracy as results of poor data quality.

Maintaining and improving data quality using appropriate methods and controls is referred to as data quality management (Gudivada et al., 2017). The implementation of data quality management improves the chances that a functional AI-based system will be developed, which will therefore successfully pass a test. Data quality management can be implemented, for instance, using company- and industry-wide data standards or standard processes to eliminate incongruities such as inconsistencies, inaccuracies or outliers. Methods for implementing data quality management are considered in detail in (Burkov, 2020; Gudivada et al., 2017).

When developing the ship recognition component, the manufacturer should use a variety of data sets to train the AI model to obtain a sufficiently large, holistic data set. The manufacturer should check the merged data set for various factors and then pose the following questions:

- Is the statistical distribution of the image data for objects in the data set for which the bearing is being calculated representative of the real-world application?
- Does it contain only image data that is consistent with the operational design domain (including clear visibility in daylight)?
- Has the image data been classified correctly?

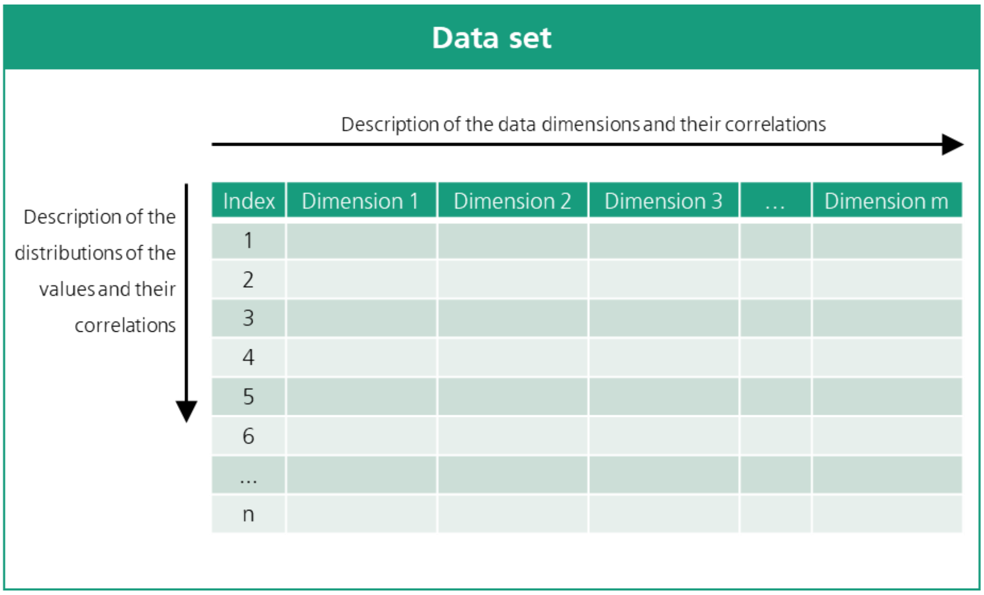
Bearing in mind the above questions, the manufacturer adjusts the data set if appropriate.

D2 | Description of the data set

The manufacturer uses one or more data sets when developing an AI-based system. The manufacturer should already have sorted the data sets being used by implementing data quality management, especially when training ML-based AI models. Because, as mentioned above, data quality, as well as the training data set used, have a significant impact on the performance of the AI models.

For the audit authority to be able to verify the performance of AI models, it must have sufficient knowledge of the properties of the training data sets that it can obtain suitable test data sets itself independently. The manufacturer should communicate these properties of the data sets to the audit authority at an early stage as part of the data description. In particular, the manufacturer must indicate if the training is taking place on publicly or commercially available data sets.

The data description should provide a description of the data sets used in development one component at a time. One way to describe a data set is illustrated below along two axes of a matrix (see Figure 20) and briefly explained using an AIS data set as an example. Data in a data set consists of one or more dimensions (horizontal row in Figure 20). The dimensions represent static and dynamic variables in an AIS data set, such as the ships' MMSI numbers, their positions and the time stamp for the data points. Each dimension can for instance be distinguished by dimension type (nominal, ordinal or numeric) and, for numerical dimensions, a distinction can also be made between discrete and continuous dimensions (Navlani et al., 2021). The possible values for each dimension can be limited by specifying the possible range of values. The relationships between dimensions must also be specified where these exist.


Data set
Description of the data dimensions and their correlations
Description of the distributions of the values and their correlations

Index	Dimension 1	Dimension 2	Dimension 3	…	Dimension m
1					
2					
3					
4					
5					
6					
…					
n					

Figure 20: Example illustrating data description using a matrix.

The various data in a data set are listed along the vertical columns. In the example with the AIS data set, AIS position messages would be listed one below the other. A suitable statistical distribution can be specified to describe the values observed for each dimension, or a sample set can be provided for estimation purposes. If possible, the relationship between the values should also be described, e.g., in time sequences.

For other techniques that can be applied when describing a data set, please refer to descriptive statistics (Navlani et al., 2021).

The data description should be submitted to the audit authority together with sample data. How the audit authority handles the data description and whether it is sufficiently detailed is explained in section 6.1 in subsection V2.

The manufacturer should describe each data set used, which the audit authority must reproduce. For image data used for ship recognition, the manufacturer has an image data set based on Figure 20 and structured as follows.

Each dimension represents a unique image. In the same row, the image content is described using additional dimensions and semantics. Additional dimensions might include weather, water and light conditions. Another dimension contains the image position data for ships in the vicinity and a corresponding dimension contains the ship type names. Each row represents one image and the associated description.

The manufacturer draws conclusions about the statistical distribution of the image content by attempting to describe the statistical distribution of the individual dimensions. One point of this attempts to answer the following question:

- Which ship types crop up, and how often?
- How many images show no ships at all?
- Do certain ship types crop up more or less frequently under certain conditions?

- Which weather, water and light conditions prevail, and how often?

The manufacturer works in a similar fashion for all the other data sets used for the remaining AI components. Descriptions of the data sets are submitted to the audit authority as part of the preliminary audit communication.

D3 | Description of model functioning based on IPO pattern

The manufacturer should provide a description of how the AI components function within the model, so that the audit authority can investigate whether the existing AI-based system is functioning (sufficiently) correctly in accordance with the manufacturer's specifications. The description of how the model functions can follow the IPO principle outlined above (see section 6.1 subsection V2).

This description should make it clear which input values should lead to which output values. This description should be sufficiently accurate so that the audit authority can verify how the model functions in the operational design domain as described, and measure and evaluate the accuracy one component-by-component according to the defined audit metrics.

The manufacturer should describe how the model functions for its AI components. The IPO principle is used for this, and to explain how the model functions based on the expected outputs for various inputs.

For the ship recognition component, considerations should include what inputs trigger ship recognition, and the need to recognize a ship. It follows from the product data sheet (see section 4.3), for instance, that the image-based bearing sensor and therefore also ship recognition only work in clear visibility in daylight. The manufacturer describes this environment as the basic requirements for the model to function properly.

Assuming the basic requirements are met, there are two cases that adequately describe how the model performs. In the first type of input, no ship of sufficient size appears in an image. Here, the ship recognition must not identify a ship anywhere in the output. In the second case, one or more ships are of sufficient size are visible. Each separate ship must be tagged in the output.

From this description of how the model performs for the ship recognition component, the audit authority can see which output is to be expected for which input. Based on this, the AI components can be tested for functionality.

Descriptions of how the model performs for each AI component are submitted to the audit authority as part of the preliminary audit communication.

8. Summary and Recommended Points of Action

The result of this study is a framework for verifying and certifying AI-based systems. This section provides a summary of the most important interim results and resulting recommended points of action for the BSH (highlighted in blue).

Putting the verification and certification procedures in a separate module K

Investigating the existing conformity assessment procedure for maritime equipment reveals that there are no applicable procedures for verifying and certifying AI-based systems (see section 3). Although the European Commission is making efforts to approve AI-based systems (see the EU AI Act in section 3.5), no legally binding or generally applicably procedures have yet been put forward.

> Considering existing publications from the European Commission and procedures in the MED, introducing a separate module into the current test and certification procedure is recommended. It makes sense to put the verification and certification procedures for AI-based systems developed in this study into this separate module (see Module K in section 3.4). This enables the proposed verification procedure to be included, without having to adapt the existing modules.

Standardizing information exchange between AI-based systems

The market analysis of existing and in-development AI-based products in the maritime environment (see section 4) results in an list of data sources (see Table 5 in Appendix A.1) and use cases (see Table 6 in Appendix A.1) for the products. Much use of camera systems has been identified (see section 4.1.1). These require more detailed consideration as there is currently no standard format for information exchange in the maritime setting.

> When testing and certifying AI-based systems, it is advisable to promote a standard format for information exchange and data sources. Such standard formats can significantly simplify and scale the testing process.

Introduction of a model-agnostic testing procedure

In order to integrate Module K into the existing testing process, this study presents a way to organize the interaction between manufacturer and audit authority, which serves as the basis for the safety and Verification Guidelines (see section 5). This process of interaction describes the framework of communication between manufacturer (Safety Guideline) and audit authority (Verification Guideline).

The entire testing process is an iterative process for testing an AI-based system step-by-step. The test is based on an evidence-based review of input and output data, applying the IPO pattern (see section 6.1 subsection V2). This procedure can be used equally well for both sAI and CI systems. The Verification Guideline is designed to be model-agnostic and makes it possible to test proper function without having to understand the functionality or architecture of the AI-based system. Breaking down the verification procedure also makes it possible to verify an AI-based system step by step and so integrate new procedures into the existing verification and certification system.

To be able to test the large number of different AI-based systems in standard way, and to ensure the future viability of the audit process, establishing a model-agnostic verification process is recommended. The verification should focus on determining “whether” rather than “how” an AI-based system works.

Formalizing the operational design domains for AI-based systems

The plan proposed considers testing the proper functioning of an AI-based system on a formalized operational design domain. Consequently, formalizing the input and output data is proposed (see section 6.1 subsections V2 and V3). This can be achieved uniformly using a method of standardized domain description, such as the operational envelope. Based on this, the audit authority can at any time determine which test data is required for the current application during a test and whether this is readily available. This plan also sets out options for measuring (see audit metrics in section 6.1 subsection V4) and evaluating (see success criteria in section 6.1 subsection V4) the proper functioning of an AI-based system (see section 6.1 subsection V4). These also enable expectations for the functionality of an AI-based system to be communicated from the audit authority to the manufacturer, as well as comparison with similar AI-based systems.

To enable standard verification, comparison with similar products and scalability of the audit processes, the use of standardized methods for formalization is recommended, both when describing the operational design domain and when measuring and evaluating functionality.

Building an automated data processing infrastructure

To scale and rapidly reproduce the verification process, development of an automated data processing infrastructure is recommended. This infrastructure should be able to obtain test data for (standardized) operational design domains, have it processed by the AI model and finally return a result measured and evaluated in accordance with set audit metrics and success criteria. Acquisition of test data, which represents the first step in such a data processing infrastructure, can benefit immensely from the use of formalized operational design domains and standardized domain descriptions (see Operational Envelope in section 6.1 subsection V3). In addition, it is shown that generating data (both through data augmentation and through data synthesis) is a promising route for targeted data procurement (see section 6.2 subsection H2). Standardized domain descriptions for the AI-based system and it being possible to generate data at any time also avoid creating unwanted data dumps.

Technical implementation of the verification procedures should be carried out using a data processing infrastructure which can be automated, to ensure scalability and reproducibility. To this end, it is of fundamental importance to rely on standardized methods of domain description for the AI products to automate the data acquisition process. Furthermore, the use of synthetic or augmented data is a promising way to obtain the necessary test data independently at any time without creating data dumps in the long term. A further advantage of using synthetic (or augmented) test data is that the audit authority can generate data that was not used by the manufacturer during development.

Considering the variability of AI-based systems and their operational design domains

This study only considers frozen AI models (see section 2.1.1), meaning those where the performance of the model does not change through interaction with the environment. It is still possible that the AI-based system will perform in an unexpected, non-tested way in time (see drift in section 2.1.1), not only due to the simple evolution of AI-based systems (such as through updates) but also due to changes in the operational design domain for the AI-based systems (relating to ship recognition systems, like when the appearance of ships evolves). Therefore, the requirements for a re-audit are set out (see section 6.3). Regarding the modularization of an AI-based system into components as described, the conditions exist to develop AI-based systems further module by module, and to integrate these changes into a previously tested AI-based system without having to go through the entire testing process on the entire system all over again.

Finally, it is recommended that the BSH at first implements the proposed model-agnostic verification and Safety Guideline into verification procedures for a simple operational design domain. Introducing measurement and performance standards as soon as possible is recommended, so that the verification procedure can be scaled and implemented right from an early stage. This gives manufacturers an expectation and offers comparability between AI-based systems within an operational design domain. Finally, it must be emphasized that it can be expected that a suitable data processing infrastructure will in particular significantly improve the feasibility, scalability and future viability of the verification procedures.

9. Acknowledgments and Final Remarks

This study was funded by the German Federal Ministry of Digital and Transport (BMDV) in the context of the BMDV Network of Experts.

The findings of this study represent the current state of knowledge of the authors at Fraunhofer CML and do not represent the views of the BSH.

The findings and considerations are based on the status of research as of August 2022.

10. List of References

ABB. (2018). *ABB Ability TM Marine Pilot Vision Modular Situational Awareness Platform*. ABB. Abgerufen am 09/09/2022, von https://library.e.abb.com/public/12ae485d68b0428884dcd453baf0c296/3AFV6116339_A_en_Pilot_Vision Leaflet.pdf

Alphabet Inc.. *Google Patents*. Google. Abgerufen am 22/06/2022, von https://patents.google.com/?q=(%22Autonomous%22+AND+%22Ship%22)&before=priority:20220101&after=priority:19900101

ASAM. (2021). *ASAM OpenODD: Concept Paper Version 1.0*. ASAM. Abgerufen am 09/09/2022, von https://www.asam.net/index.php?eID=dumpFile&t=f&f=4544&token=1260ce1c4f0afdbe18261f7137c689b1d9c27576

Avikus. (a) *Avikus AiBOAT*. Avikus. Abgerufen am 09/09/2022, von https://www.avikus.ai/eng/product/aiboat

Avikus. (b) *Avikus HiNAS*. Avikus. Abgerufen am 09/09/2022, von https://www.avikus.ai/eng/product/hinas

Brooks, S. K., & Greenberg, N. (2022). Mental health and psychological wellbeing of maritime personnel: a systematic review. *BMC Psychology*, *10*(1), 1–26. https://doi.org/10.1186/s40359-022-00850-4

BSB AI. (2022). *Oscar Navigation Products*. BSB AI. Abgerufen am 09/09/2022, von https://www.oscar-navigation.com/#products

Bundesamt für Schifffahrt und Hydrographie. (2022). *Nationale Zulassung*. BSH. Abgerufen am 09/09/2022, von https://www.bsh.de/DE/THEMEN/Schifffahrt/Schiffsausruestung_Marktueberwachung/Nationale_Zulassung/nationale_zulassung_node.html

Burkov, A. (2019). *The Hundred-Page Machine Learning Book*. Andriy Burkov.

Burkov, A. (2020). *Machine Learning Engineering*. True Positive Inc.

Burmeister, H. C., Constapel, M., Ugé, C., & Jahn, C. (2020). From Sensors to MASS: Digital Representation of the Perceived Environment enabling Ship Navigation. *IOP Conference Series: Materials Science and Engineering*, *929*(1). https://doi.org/10.1088/1757-899X/929/1/012028

Captain AI. (2022). *Technology – Captain AI*. Captain AI. Abgerufen am 09/09/2022, von https://www.captainai.com/technology/

Danish Maritime Authority. (2018). *Analysis of regulatory barriers to the use of autonomous ships. Final Report*. https://dma.dk/Media/637745499808186153/Analysis of Regulatory Barriers to the Use of Autonomous Ships.pdf

Daranda, A., & Dzemyda, G. (2020). Navigation decision support: Discover of vessel traffic anomaly according to the historic marine data. *International Journal of Computers, Communications and Control*, *15*(3), 1–9. https://doi.org/10.15837/IJCCC.2020.3.3864

DIN. (2011). *DIN EN 61162-1:2011-09 Navigations- und Funkkommunikationsgeräte und -systeme für die Seeschifffahrt - Digitale Schnittstellen - Teil 1: Ein Datensender und mehrere Datenempfänger*. https://www.beuth.de/de/norm/din-en-61162-1/143482391

DIN. (2020). DIN SPEC 13266:2020-04 Leitfaden für die Entwicklung von Deep-Learning-Bilderkennungssystemen. In *DIN SPEC (PAS)* (Ausgabe April 2020). https://doi.org/https://dx.doi.org/10.31030/3134557

Ekbatani, H. K., Pujol, O., & Segui, S. (2017). Synthetic Data Generation for Deep Learning in Counting Pedestrians. *Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods*, *2017-Januar*, 318–323. https://doi.org/10.5220/0006119203180323

Etzkorn, P. (2022). Der Schiffszusammenstoß unter Beteiligung autonom fahrender Schiffe. In *Der Schiffszusammenstoß unter Beteiligung autonom fahrender Schiffe*. Nomos Verlagsgesellschaft mbH & Co. KG. https://doi.org/10.5771/9783748934073

Europäische Kommission. (2021). *Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für Künstliche Intelligenz (Gesetz über Künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union*. Europäische Kommission. https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A52021PC0206

Europäische Kommission. (2022). *Durchführungsverordnung (EU) 2022/1157 der Kommission vom 4. Juli 2022 mit Vorschriften für die Anwendung der Richtlinie 2014/90/EU des Europäischen Parlaments und des Rates hinsichtlich der Entwurfs-, Bau- und Leistungsanforderungen sowie der Prüfnormen* (s. 1–243). Amtsblatt der Europäischen Union. http://data.europa.eu/eli/reg_impl/2022/1157/oj

Europäisches Parlament und Rat der Europäischen Union. (2014). *Richtlinie 2014/90/EU des europäischen Parlaments und des Rates vom 23. Juli 2014 über Schiffsausrüstung und zur Aufhebung der Richtlinie 96/98/EG des Rates (2014/90/EU)* (s. 146–185). Amtsblatt der Europäischen Union. http://data.europa.eu/eli/dir/2014/90/oj

Flasiński, M. (2016). *Introduction to Artificial Intelligence*. Springer International Publishing. https://doi.org/10.1007/978-3-319-40022-8

Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J., & Greenspan, H. (2018). Synthetic data augmentation using GAN for improved liver lesion classification. *Proceedings - International Symposium on Biomedical Imaging*, *April 2018*, 289–293. https://doi.org/10.1109/ISBI.2018.8363576

Gudivada, V. N., Ding, J., & Apon, A. (2017). Data Quality Considerations for Big Data and Machine Learning: Going Beyond Data Cleaning and Transformations. *International Journal on Advances in Software*, *10.1*, 1–20. https://www.researchgate.net/publication/318432363

Gyllenhammar, M., Johansson, R., Warg, F., Chen, D., Heyn, H.-M., Sanfridson, M., Söderberg, J., Thorsén, A., Ursing, S., Ab, Z., & Com, M. G. (2020). Towards an Operational Design Domain That Supports the Safety Argumentation of an Automated Driving System. *10th European Congress on Embedded Real Time Systems*, 1–10. https://www.diva-portal.org/smash/get/diva2:1390550/FULLTEXT01.pdf

IHO. (2014). *Specifications for Chart Content and Display Aspects of ECDIS*. Abgerufen am 10/04/2022, von www.iho.int

IMO. (1971). *Resolution A.224(VII), Performance standards for Echo-Sounding equipment*. https://wwwcdn.imo.org/localresources/en/KnowledgeCentre/IndexofIMOResolutions/AssemblyDocuments/A.224(7).pdf

IMO. (1995). *Resolution A.819(19), Recommendation on Performance Standards for*

Shipborne Global Positioning System (GPS) Receiver. https://wwwcdn.imo.org/localresources/en/KnowledgeCentre/IndexofIMOResolutions/AssemblyDocuments/A.819(19).pdf

IMO. (1998). *Resolution MSC.74(69), Adoption of New and Amended Performance Standards*. https://wwwcdn.imo.org/localresources/en/OurWork/Safety/Documents/AIS/Resolution MSC.74(69).pdf

IMO. (2001). *Resolution A.915(22), Revised maritime policy and requirements for a future Global Navigation Satellite System (GNSS)*. https://wwwcdn.imo.org/localresources/en/KnowledgeCentre/IndexofIMOResolutions/AssemblyDocuments/A.915(22).pdf

IMO. (2004). *Resolution MSC.192(79), Adoption of the Revised Performance Standards for Radar Equipment*. https://wwwcdn.imo.org/localresources/en/KnowledgeCentre/IndexofIMOResolutions/MSCResolutions/MSC.192(79).pdf

IMO. (2015). *Resolution A.1106(29), Revised guidelines for the onboard operational use of shipborne automatic identification systems (AIS)*. https://wwwcdn.imo.org/localresources/en/OurWork/Safety/Documents/AIS/Resolution A.1106(29).pdf

IMO. (2017). *MSC.1/Circular.1575, Guidelines for Shipborne Position, Navigation And Timing (PNT) Data Processing*. https://www.imorules.com/MSCCIRC_1575.html

IMO. (2022). *Maritime Safety Committee (MSC 105), 20-29 April 2022*. International Maritime Organization. Abgerufen am 24/04/2022, von https://www.imo.org/en/MediaCentre/MeetingSummaries/Pages/MSC-105th-session.aspx

ITU. (2014). *Technical characteristics for an automatic identification system using time division multiple access in the VHF maritime mobile frequency band (Recommendation ITU-R M.1371-5)*. https://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.1371-5-201402-I!!PDF-E.pdf

Kongsberg. (2017). *Autonomy is here - Powered by Kongsberg*. Kongsberg. Abgerufen am 09/09/2022, von https://www.kongsberg.com/maritime/about-us/news-and-media/our-stories/autonomy-is-here--powered-by-kongsberg/

Kongsberg. (2019). *Kongsberg Autonomous Shipping*. Kongsberg. Abgerufen am 09/09/2022, von https://www.kongsberg.com/maritime/support/themes/autonomous-shipping/?_t_id=MJquBrAbUl9fFaQl-KpGfQ%3D%3D&_t_uuid=fbm8o_i1TbajRZ0iocOo4w&_t_q=Autonomous+shipping&_t_tags=language%3Aen%2Csiteid%3A24c9be7d-c7a0-47ff-9aff-d09ef8b15bbc%2Candquerymatch&_t_hit.

Kongsberg. (2022). *Kongsberg Situation Awareness*. Kongsberg. Abgerufen am 09/09/2022, von https://www.kongsberg.com/maritime/products/situational-awareness/?_t_id=l_tj8xlY0kjRMfKEZQECPA%3D%3D&_t_uuid=sGRclj3wQHygY_d6z1awwg&_t_q=Intelligent+Awareness&_t_tags=language%3Aen%2Csiteid%3A24c9be7d-c7a0-47ff-9aff-d09ef8b15bbc%2Candquerymatch&_t_hit.id

Korakakis, M., Mylonas, P., & Spyrou, E. (2018). A short survey on modern virtual environments that utilize AI and synthetic data. *MCIS 2018 Proceedings*, 34. https://aisel.aisnet.org/mcis2018/34

Marine AI Ltd. (2022). *Guardian by Marine AI*. Abgerufen am 12/09/2022, von https://marineai.co.uk/products/guardian/#autonomy

Mayflower Autonomous Ship. (2022). *Mayflower Autonomous Ship - Technology*. Abgerufen am 09/09/2022, von https://mas400.com/technology

Minter, A. (2021). *The Next Shipping Crisis: A Maritime Labor Shortage*. Bloomberg.Com. Abgerufen am 26/09/2022, von https://www.bloomberg.com/opinion/articles/2021-11-06/the-next-shipping-crisis-a-maritime-labor-shortage

Navlani, A., Fandango, A., & Idris, I. (2021). *Python Data Analysis: Perform data collection, data processing, wrangling, visualization, and model building using Python* (3rd ed.). Packt Publishing.

Nikolenko, S. I. (2021a). *Synthetic Data for Deep Learning* (Vol. 174). Springer International Publishing. https://doi.org/10.1007/978-3-030-75178-4

Nikolenko, S. I. (2021b). Introduction: The Data Problem. In *Springer Optimization and Its Applications* (Vol. 174, WP. 1–17). Springer International Publishing. https://doi.org/10.1007/978-3-030-75178-4_1

Norvig, P., & Russell, S. J. (2021). *Artificial Intelligence: a modern approach* (4th ed.). Pearson.

Orca AI. (2022). *Orca AI - Solutions*. Orca AI. Abgerufen am 09/09/2022, von https://www.orca-ai.io/solutions

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). *Hierarchical Text-Conditional Image Generation with CLIP Latents*. https://doi.org/10.48550/arXiv.2204.06125

Riveiro, M., Pallotta, G., & Vespe, M. (2018). Maritime anomaly detection: A review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(5), e1266. https://doi.org/10.1002/widm.1266

Rødseth, Ø. J., Lien Wennersberg, L. A., & Nordahl, H. (2022). Towards certification of autonomous ship systems by their operational envelope. *Journal of Marine Science and Technology*, *27*(1), S. 67-76. https://doi.org/10.1007/s00773-021-00815-z

Saildrone Inc. (2022). *Saildrone Voyager*. Abgerufen am 09/09/2022, von https://assets.website-files.com/5beaf972d32c0c1ce1fa1863/629a8318e699b0b1981d06c4_SD_Voyager-Bathymetry_Product_Card_r8-web final-2206.pdf

Samek, W., & Müller, K. (2019). Towards Explainable Artifical Intelligence. In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning* (WP. 5–22). Springer, Cham. https://doi.org/10.1007/978-3-030-28954-6_1

Samsung Heavy Industries. *SHI SVessel Onboard Solution*. Abgerufen am 12/09/2022, von https://shi.svessel.com/?page_id=298

Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, *2*(3), 160. https://doi.org/10.1007/s42979-021-00592-x

Sea Machines Robotics. (2022). *Sea Machines 300*. Abgerufen am 09/09/2022, von https://sea-machines.com/wp-content/uploads/SM-300-Insert-Sheet_2022_final-web.pdf

Seadronix. (2022). *Seadronix - Our Products*. Abgerufen am 09/09/2022, von https://www.seadronix.com/products

Seafar NV. (2022). *Seafar Services*. Abgerufen am 09/09/2022, von https://seafar.eu/services/

Seib, V., Lange, B., & Wirtz, S. (2020). *Mixing Real and Synthetic Data to Enhance Neural Network Training -- A Review of Current Approaches*. https://doi.org/10.48550/arXiv.2007.08781

Skredderberget, A. (2018). *Yara Birkeland - The first ever zero emission, autonomous ship*. Abgerufen am 09/09/2022, von https://www.yara.com/knowledge-grows/game-changer-for-the-environment/

Tsirikoglou, A., Kronander, J., Wrenninge, M., & Unger, J. (2017). *Procedural Modeling and Physically Based Rendering for Synthetic Data Generation in Automotive Applications*. http://arxiv.org/abs/1710.06270

Wang, W., Shan, T., Leoni, P., Fernandez-Gutierrez, D., Meyers, D., Ratti, C., & Rus, D. (2020). Roboat II: A Novel Autonomous Surface Vessel for Urban Environments. *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 1740–1747. https://doi.org/10.1109/IROS45743.2020.9340712

Wärtsilä. (2022). *Wärtsilä Advanced Assistance Systems*. Abgerufen am 09/09/2022, von https://www.wartsila.com/voyage/autonomy-solutions/advanced-assistance-systems

Weller, A. (2019). Transparency: Motivations and Challenges. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 11700, Issue Section 2, WP. 23–40). Springer Cham. https://doi.org/10.1007/978-3-030-28954-6_2

Yoshida, M., Shimizu, E., Sugomori, M., & Umeda, A. (2021). Identification of the Relationship between Maritime Autonomous Surface Ships and the Operator's Mental Workload. *Applied Sciences*, *11*(5), 2331. https://doi.org/10.3390/app11052331

Zhang, C., Kuppannagari, S. R., Kannan, R., & Prasanna, V. K. (2018). Generative Adversarial Network for Synthetic Time Series Data Generation in Smart Grids. *2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, 1–6. https://doi.org/10.1109/SmartGridComm.2018.8587464

Žliobaitė, I. (2010). *Learning under Concept Drift: an Overview*. https://doi.org/10.48550/arXiv.1010.4784

A Appendix

A.1. Results of the Market Analysis

Table 5: Companies or products reviewed in the market analysis and their data sources.

Company/product	Sources	RGB camera	Infra-red camera	LIDAR/RADAR	IMU/MRU	GNSS	AIS	Weather data/ sensors	Depth sounding
ABB Ability Marine Pilot Vision	(ABB, 2018)	x	x	x	x	x	x		
Avikus AiBOAT	(Avikus, a)	x		x		x	x		
Avikus HiNAS	(Avikus, b)	x	x	x		x	x		
Captain AI	(Captain AI, 2022)	x		x		x	x		
Kongsberg Situation Awareness	(Kongsberg, 2022)	x		x	x	x	x		
Kongsberg Maritime Autonomous Shipping	(Kongsberg, 2019)	x		x	x	x	x		
Marine AI Guardian Autonomy	(Marine AI Ltd, 2022)	x		x			x	x	
Mayflower Autonomous Ship[1]	(Mayflower Autonomous Ship, 2022)	x		x	x	x	x	x	x
Orca AI	(Orca AI, 2022)	x	x						
Oscar	(BSB AI, 2022)	x	x		x	x			
Roboat	(Wang et al., 2020)	x		x	x	x			
Saildrone	(Saildrone Inc., 2022)	x			x	x	x	x	x
Sea Machines SM300	(Sea Machines Robotics, 2022)	x		x	x	x	x		x
Seadronix AVISS	(Seadronix, 2022)	x							
Seafar	(Seafar NV, 2022)	x		x	x	x	x		
SVessel Samsung Heavy Industries	(Samsung Heavy Industries)	x					x		
Wärtsilä Voyage Autonomy Solutions	(Wärtsilä, 2022)	x	x	x		x	x	x	x
Yara Birkeland[1]	(Kongsberg, 2017; Skredderberget, 2018)	x	x	x	x	x	x		

[1] Combines AI-based products as a system integrator.

Table 6: Companies or products reviewed in the market analysis and their applications.

Company/product	COLREGs evaluation	Collision avoidance	Obstacle and coast recognition	Ship recognition	Route planning	Docking and cast-off support
ABB Ability Marine Pilot Vision	x		x	x		
Avikus AiBOAT	x	x	x	x	x	x
Avikus HiNAS				x	x	
Captain AI		x		x	x	
Kongsberg Situation Awareness			x	x		x
Kongsberg Maritime Autonomous Shipping	x	x	x	x	x	
Marine AI Guardian Autonomy		x	x	x	x	
Mayflower Autonomous Ship[1]	x	x	x	x	x	
Orca AI		x	x	x		
Oscar	x	x	x	x		
Roboat		x	x	x		
Saildrone	x	x	x	x	x	
Sea Machines SM300	x	x	x	x		
Seadronix AVISS			x	x		x
Seafar	x	x	x	x	x	
SVessel Samsung Heavy Industries				x		
Wärtsilä Voyage Autonomy Solutions			x			x
Yara Birkeland[1]	x	x	x	x	x	x

[1] Combines AI-based products as a system integrator.