STUDY OF MARITIME AUTONOMOUS SURFACE SHIPS (MASS) TRUSTWORTHINESS: HARDWARE POINT OF VIEW

Hosna Namazi
UiT, The Arctic University of Norway
Tromsø, Norway

Lokukaluge Prasad Perera
UiT, The Arctic University of Norway, Tromsø, Norway
SINTEF Digital, SINTEF AS, Oslo, Norway

ABSTRACT

Nowadays, the maritime industry, like other industries, is incorporating Machine Learning (ML) and Artificial Intelligence (AI) approaches in their applications. Since the rise of Maritime Autonomous Surface Ships (MASS) is on the horizon, such intelligent algorithms would replace conventional ship navigation with a higher level of autonomy. In other words, a digital navigator can be developed based on the data obtained from the human navigator's actions when controlling vessels. To ensure the prosperity of these vessels, the trustworthiness of such navigation actions must be guaranteed. Generally, the trustworthiness of any AI-based application can be studied from two primary levels: software and hardware. The software algorithms of trustworthy digital navigators should be Explainable, Fair, and Responsible. Besides, two concepts of Resilience and Availability must be confirmed for the hardware used for their development. Although the trustworthiness of the AI-based application from the software level is mainly focused on the previous research study, the trustworthiness of the hardware level should not be neglected. This preliminary study looks into ship systems used in such applications and then focuses on the digital navigator's trustworthiness at a hardware level. It identifies the most appropriate key performance indicators for studying this topic, and proper approaches to investigate them are summarized from the literature.

KEYWORDS: MASS; Digital Navigator; Hardware Trustworthiness; Artificial Intelligence; Machine Learning.

INTRODUCTION

In recent years, the world has witnessed unprecedented development in implementing Artificial Intelligence (AI) and Machine Learning (ML) methods, i.e., driverless vehicles, ChatGPT, etc., into different industrial applications, with a large number of research studies being done. This is due to the recent development in the field of sensor technology, data, and ML and AI algorithms. Like any other field, the maritime industry has realized the advantages of exploiting these new methods for some time now and has invested in studying the potential of these algorithms in improving energy efficiency of ship navigation and making the advent of Maritime Autonomous Surface Ships (MASS) a reality. As an example, one of the shipping industry's main focuses is implementing such methods to develop digital navigators in autonomous ships.

According to the International Maritime Organization (IMO), the MASS concept is defined as a vessel type that must have the ability to navigate safely and efficiently without human intervention to a varying degree (IMO, 2018). These vessel types are categorized into four degrees based on their autonomy level as follows: i) degree one is dedicated to the ships with automated processes and decision support with seafarers onboard, ii) degree two is remotely controlled ships with seafarers onboard, iii) degree three is remotely controlled ships without seafarers onboard, and iv) degree four is a fully autonomous ship. Based on this categorization, it is evident that as the autonomy increases, the ships need to be equipped with decision support facilities that assist them in making proper decisions.

Autonomous and remotely controlled ships can bring along many benefits to the shipping industry, such as economic advantages due to their lower operational and maintenance costs, reduced fuel consumption, and, consequently, guiding towards a greener industry (Vartdal et al., 2018). In addition, MASS can reduce human-related casualties and crises and help with crew shortages since the number of humans on board is limited in some situations (Kim et al., 2022). More importantly, various researchers have proclaimed human errors to be the primary cause of accidents, with fatigue, stress, and overwhelming workload as the main reasons (Apostol-mates & Barbu, 2016). This can be overcome by introducing an appropriate human-machine system interface in MASS due to its superiority in handling large amounts of data and information in opposition to the limited capability of human minds (Huang et al., 2020). As mentioned, MASS is supposed to make decisions without human intervention with the help of support algorithms online. However, classic mathematical models usually lack the capability to handle large-scale data sets and make decisions in real-time; thus, powerful data analysis tools are needed in these types of situations. As an example, different data-driven methods, such as clustering with Gaussian Mixture Models, can be used for finding different operating regions of vessels using recorded data onboard, which can be utilized towards developing localized models for the respective vessels as a part of the industrial digitalization in shipping (Taghavi & Perera, 2022 & 2023). Countless studies have been dedicated to creating models that can support MASS in complex navigational environments. Deep-learning-based methods have gained significant attention for developing collision avoidance systems (Perera, 2020). Murray & Perera (2021) developed a deep-learning model that can predict the ship routes for the future 30 minutes using Automatic Identification System (AIS) data.

Although the emergence of MASS has the potential to solve issues associated with conventional ships, it can also introduce new challenges to the industry. The role of ship navigation that was previously part of human operators' duties is now going to be handed out to the algorithms that run themselves. Thus, the transition of knowledge and experience from human navigators to digital navigators requires considerable resources and expertise, such as changes and integrations that need to be done to the existing systems as well as the rules and regulations of ship navigation.

One of the downsides of this transition is the limitations of algorithms like neural networks in handling complex tasks according to the rules of the sea, whereas humans can manage these situations with training and education. In other words, human minds are superior to machines in perception and judgment in some situations. A prominent example of this is when an ethically challenging situation arises. In such conditions, human operators usually can take the initiative and show the best performance based on their knowledge and innate ability gained through training and experience. However, the algorithms behind digital navigators, or any other AI-based applications, are programmed to act in a specific way in a specific situation, and there will be a lack of improvisation in such complex scenarios. This is where the importance of trustworthiness studies gets highlighted. In other words, since algorithms are going to take humans' place in shipping, they should prove that they have similar safe navigation abilities as humans when dilemmas happen.

Before identifying a framework for the trustworthiness evaluation of a digital navigator, it is better to take a step back to see how a digital navigator can be developed and how it works. The following section will address that topic.

THE DIGITAL NAVIGATOR CONCEPT

As mentioned before, one of the AI applications that currently has gotten attention in the maritime field is the development of digital navigators, which have the same role as the human officer in a manned ship. Digital navigators are expected to possess capabilities similar to humans in navigating ships in case of dilemmas since they will take the human role in ship navigation. Therefore, the general idea for developing this new type of navigator is to clone human navigators' skills, knowledge, and experience by the data collected from real scenarios. Then, these data sets are used to train deep neural networks. It is worth reminding that the goal here is to extract the pattern of human skills, knowledge, and experience and train the networks to learn how humans improvise. Of course, to prevent learning erroneous human behaviors, multiple data sets should be used for training. These data sets contain the right decisions, which must comply with the regulations, and the wrong decisions will be detected by AI-based methods as anomalies, removed, if possible recovered, and even compensated with the correct data.

As shown in Fig. 1, the key pillars of successful ship navigation are complying with existing sea rules and regulations and the ability to interact with navigation and automation systems and technologies, which must be considered in the design of digital navigators. Moreover, intelligent ships should be able to interact with humans on manned vessels since there will be a time when both types of ships, along with remotely controlled ones will coexist in the sea environment. This future sea environment is also called a mixed navigational environment. The following paragraphs present a detailed description of the digital navigator's decision-making process.

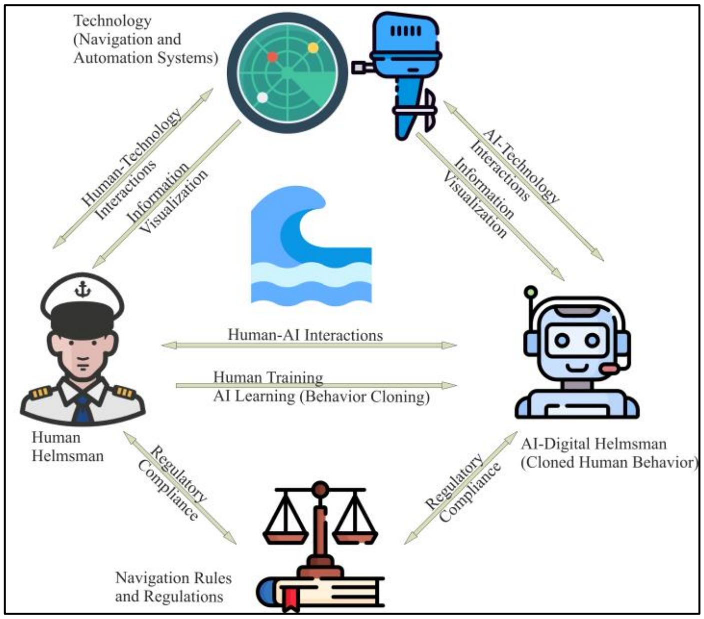

Fig 1. Key pillars in the development of digital navigators (Namazi & Perera, 2023)

For a successful implementation of digital navigators, various supporting services and authorities are needed to provide the required information to autonomous ships prior to making decisions. Maritime and port authorities, weather centers, vessel traffic managemen and information systems are a number of information sources that can provide energy efficiency and emission control rules and regulations, digital maps and navigation regulations, weather information, and vessel traffic information, respectively. This information will aid the decision-support layers of digital navigators in voyage and route planning, ship stability calculations, and collision avoidance systems, to name a few.

On the other hand, hardware components used in navigation systems, including radar and automatic radar plotting aids (ARPAa), Electronic Chart Display and Information Systems (ECDIS), as well as onboard IoT systems such as Global navigation satellite system (GNSS), Light Detection and Ranging (LiDAR), weather transmitters provide the real-time navigation information and data to the digital navigator and its support layers. The digital navigator, which is developed based on cloning human navigators' skills, makes the proper navigation decisions that are sent to ship control systems such as propeller and rudder control systems. In Fig. 2, a schematic diagram of the decision-making process of digital navigators is illustrated. By understanding the role of digital navigators in future autonomous ships, the importance of introducing a trustworthiness framework is now highlighted in this study as the main contribution.

AI TRUSTWORTHINESS EVALUATION FRAMEWORK

Now that the definition of a digital navigator is discussed extensively, the trustworthiness of such concepts can be addressed. Any trustworthy AI-based application, including digital navigators, must have some mutual characteristics. Since software and hardware systems are integrated for the development of digital navigators, it is reasonable to do the trustworthiness study from both levels. The algorithms used to develop digital navigators can be trusted if they are explainable, fair, and accountable. Also, reliability, security, privacy, and safety are essential features for both software and hardware systems incorporated in a trustworthy AI application. A trustworthiness study framework for the software level of digital navigators has been introduced by Namazi & Perera (2023). Fig. 3 displays this evaluation framework for both software and hardware levels. In the following, a summary of the main characteristics of the algorithms of a trustworthy AI application is given.

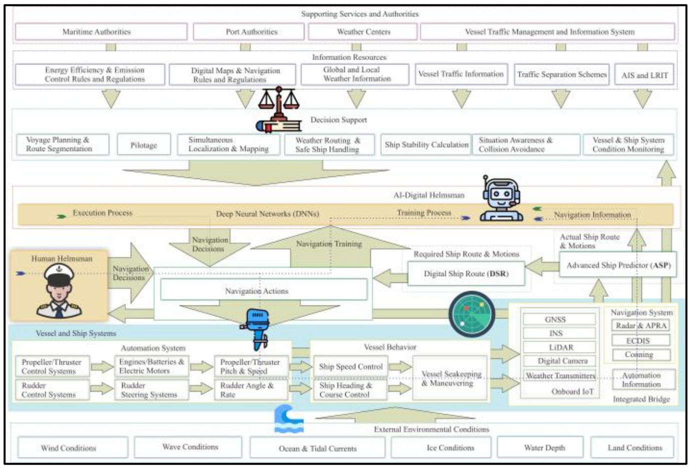
Supporting Services and Authorities
Maritime Authorities
Port Authorities
Weather Centers
Vessel Traffic Management and Information System
Information Resources
Energy Efficiency & Emission Control Rules and Regulations
Digital Maps & Navigation Rules and Regulations
Global and Local Weather Information
Vessel Traffic Information
Traffic Separation Schemes
AIS and LRIT
Decision Support
Voyage Planning & Route Segmentation
Pilotage
Simultaneous Localization & Mapping
Weather Routing & Safe Ship Handling
Ship Stability Calculation
Situation Awareness & Collision Avoidance
Vessel & Ship System Condition Monitoring
AI-Digital Helmsman
Execution Process
Deep Neural Networks (DNNs)
Training Process
Navigation Information
Navigation Decisions
Navigation Training
Actual Ship Route & Motions
Human Helmsman
Navigation Decisions
Required Ship Route & Motions
Advanced Ship Predictor (ASP)
Digital Ship Route (DSR)
Navigation Actions
Vessel and Ship Systems
GNSS
INS
LiDAR
Digital Camera
Weather Transmitters
Onboard IoT
Navigation System
Radar & APRA
ECDIS
Conning
Automation Information
Integrated Bridge
Automation System
Propeller/Thruster Control Systems
Engines/Batteries & Electric Motors
Propeller/Thruster Pitch & Speed
Rudder Control Systems
Rudder Steering Systems
Rudder Angle & Rate
Vessel Behavior
Ship Speed Control
Ship Heading & Course Control
Vessel Seakeeping & Maneuvering
External Environmental Conditions
Wind Conditions
Wave Conditions
Ocean & Tidal Currents
Ice Conditions
Water Depth
Land Conditions

Fig 2. A schematic diagram of the decision-making process of digital navigators

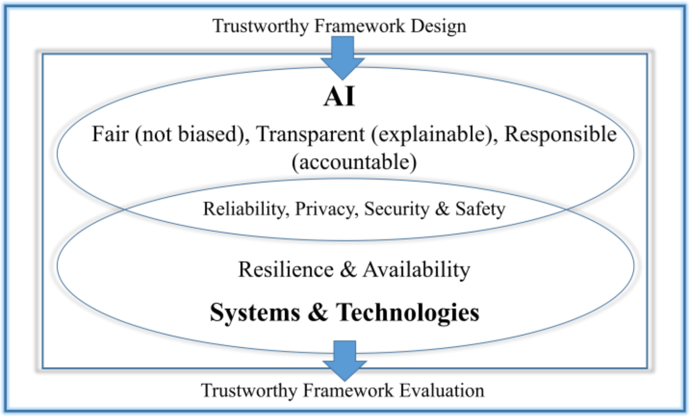
Trustworthy Framework Design
AI
Fair (not biased), Transparent (explainable), Responsible (accountable)
Reliability, Privacy, Security & Safety
Resilience & Availability
Systems & Technologies
Trustworthy Framework Evaluation

Fig 3. Trustworthiness Evaluation Framework of AI (Namazi & Perera, 2023)

Explainability/transparency assists in understanding the decision-making logic behind the algorithms. It can also help with the bias detection and debugging of the programs. Considering the complex nature of neural networks used for the development of digital navigators, such as Multi-layer Perceptron Neural Networks (MLP), Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN), some post-hoc local explanations and feature relevance techniques are suggested to be used (Montavon et al., 2017)(Barredo et al., 2020)(Lipton, 2018)(Zeiler & Fergus, 2014).

Fairness and unbiasedness are other essential features of a trustworthy application. Although AI-based systems are often susceptible to biases, they are expected to make decisions fairly and without inclinations toward a specific group. For this purpose, a framework was chosen that looks for any possible biases through every lifecycle stage of developing an application (Agarwal & Agarwal, 2022).

A **responsible AI** is defined to be functional, legal, ethical, and it must also fulfill its philanthropic responsibility. For this purpose, keen consideration must be put into defining the functional values and social issues that will be addressed by the application. Subsequently, the most proper methods and approaches for developing it must be selected. After development, validation and performance monitoring must be done (Cheng et al., 2021). Finally, it was concluded that **explainability** is the most important feature since it can be said that it is a requirement for having a fair and responsible application.

To apply the proposed software trustworthiness evaluation framework mentioned in (Namazi & Perera, 2023), and the hardware trustworthiness evaluation framework from this current paper, a research vessel has been developed by the same research group under the "UiT Autonomous Ship Program" in the Arctic University of Norway (UiT). In Fig. 4, a schematic picture of this UiT research vessel is shown. The respective hardware and software frameworks that are used in the same vessel, and their placement are also depicted in this figure. This vessel is designed to be remotely operated from the onshore operation center (OOC) as well at the initial stage of this project. To this end, several arrangements are realized. For instance, it consists of a server for handling and processing sensor data, and microprocessor-based controllers for managing the rudder & propulsion control system, and a computer for user interface and oversight.

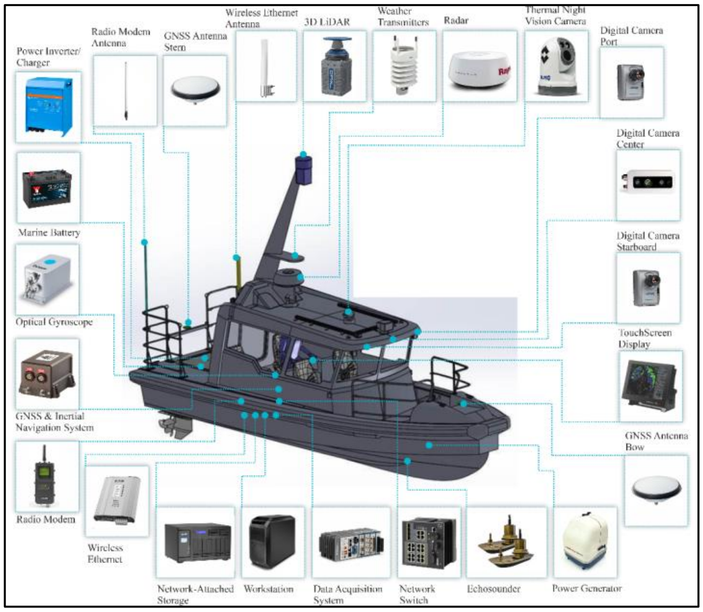

Fig 4. Schematic diagram of UiT Autonomous Ship and its hardware components

The following section addresses the necessity of hardware trustworthiness as the main contribution of this study. First, the key concepts of this study framework will be mentioned, and then the question of how these concepts can be extended to the digital navigators will be answered in further detail.

HARDWARE TRUSTWORTHINESS FRAMEWORK

Trustworthiness in system hardware means a reliable continuation of data transfer between various systems. Therefore, the ML algorithms executing in these frameworks can be supported by the trustworthiness of system hardware. In other words, such systems must adapt and tolerate vulnerabilities such as failures throughout their lifecycle (Junior & Kamienski, 2021).

As has already been mentioned, the effectiveness of AI applications does not solely depend on the software part of it, but it is also highly dependent on the hardware aspect. The hardware trustworthiness framework study is a critical concept not only in the development process of AI applications but also after it is commercialized into various industrial applications. It considers the physical side of the applications to ensure they are resilient, available, reliable, and secure at the time of their use. In the following, each of the respective key features is discussed.

Resilience

This feature is pivotal in ensuring the hardware systems can operate under a wide range of conditions. In other words, these systems must be designed to deal with errors without losing their functionality and integrity due to the fact that these applications usually work with sensitive and critical data, and their decisions are impacted by them. One scenario that is most likely to happen for AI applications is exposure to cyber threats. Making these applications resilient against cyber-attacks can be done by implementing secure boot processes, hardware-based encryption, intrusion detection systems, and robust mechanisms to recover quickly from these attacks (Dave et al., 2021).

Another feature of a resilient system is the ability to tolerate abrupt environmental changes, especially in the case of autonomous ships that are expected to work in diverse weather conditions. This can be of crucial importance since the range of temperature they work in varies a lot. Also, wind, humidity, and component corrosion due to water salinity are other environmentally challenging factors that must be considered while selecting ship systems. The use of durable materials and protective enclosures can address this issue.

As was elaborated earlier, the UiT autonomous vessel is developed to be exposed to extreme environmental conditions in terms of low temperature and humidity in the Arctic area. For this reason, various measures are being employed in the designing and installation process of hardware equipment. Selecting marine-grade systems can guarantee the proper performance of the respective equipment by having resistance to corrosion and prolonging the life span of the vessel and such systems are selected for the UiT research vessel. In addition, the vessel structure is made from the marine-grade materials that are commonly used in this field including different grades of stainless steel such as grade 316 and grade 304, alloy steel, galvanized steel, and aluminum. Each of these materials has a characteristic that makes it proper for a special environment.

Apart from the materials that are chosen according to the environment that the ship will be operating in, some other measures can be considered to increase the durability of the ship. For instance, internal heating systems are employed to maintain the proper temperature for the hardware systems. Also, some of the computers are running constantly which will help with maintaining the temperature. In case of a power failure, backup batteries are embedded in the ship to pick it up, and if shutdowns happen for a longer period of time, alarm systems will send a message and notify the responsible persons.

Another feature that can help autonomous ships to be resilient is the ability to predict component maintenance. In recent years, the idea of harnessing data from various components to predict possible failures has gained significant attention. This proactive approach enables necessary maintenance to be performed before ships embark on open-sea voyages, where assistance is scarce. Implementing additional sensors and diagnostic tools to monitor the health of these systems can help with detecting potential failures.

In a research study dedicated to this topic, Bolchini et al. (2023) offer a comprehensive review of the state-of-the-art resilience analysis and hardening methods in deep learning applications in case of hardware failures. The authors suggested two main methods for improving the resilience of such applications. The first one is redundancy-based techniques, where extra components are added to compensate if a failure happens. This technique includes methods such as Duplication with Comparison (DWC), where errors are detected by repeating and comparing the outputs; Triple Modular Redundancy (TMR), in which three components are used in parallel to find the correct output with the majority voting, and Error Correcting Codes (ECCs) that are used for detecting errors in data storage and transmission. Introducing redundancy in system components is known to be one of the primary methods of failure prevention and has been around for several decades in different industries such as the airborne industry (Hershey, 1956). Eriksen & Lützen (2022) have studied how redundancy can affect the reliability of machinery systems on unmanned ships by simulating a case study. The results show that while redundancy can help mitigate risks related to independent failures, it does not necessarily benefit the cases with dependent failures. To improve the system reliability in dependent failure cases, measures such as predictive maintenance, use of higher-quality components, and failure mode and analysis are suggested to be used.

The second method is deep learning-based techniques, in which the models are trained by data sets that include hardware faults or noise. The fault-aware method is an example of this technique and makes the trained models more robust against similar issues. These approaches increase the adaptability of the trained models, allowing them to maintain performance in the presence of hardware errors.

In the work of Wan et al. (2022), the resilience analysis and improvement methods are focused explicitly on autonomous systems. In this paper, two other approaches have been suggested aside from the previously discussed methods. The Lightweight, Application-Aware Fault Detection and Mitigation method is a proper choice for the resilience improvement of resource-constrained autonomous systems. In this method, lightweight ML algorithms are used to detect and mitigate faults in the operation time of the algorithms so that the application performance is not jeopardized. Another method mentioned in this paper is the Intelligent Fault Injection (FI) scheme. In this approach, the errors are deliberately injected into the system in the test phase to simulate the faults that can occur due to hardware failures. This method helps to identify the system’s vulnerability efficiently and to understand how well it can cope with possible failures.

Availability

Another critical concept in the hardware trustworthiness of AI-based applications is the availability of the systems during the operation phases of these applications. Availability is a measure of the system’s accessibility and operational status when required for use. In other words, it shows the reliability and uptime of the system, and it is often quantified as a percentage of uptime in a given time period. For the AI systems that are required to make real-time decisions, such as autonomous vessels, high availability of ship systems is essential, and the system's inaccessibility due to issues like maintenance, unexpected failures, or upgrades must be minimized.

An available system typically must be supported by a robust underlying network and infrastructure that is able to support the operations continuously, making it more reliable. To this end, recovery strategies are usually effective for data recovery or system restoration in case of failure. Another aspect of an available system is when the components are easy to repair or upgrade with minimum downtime. Usually, redundant key elements are embedded to ensure continuous operations.

Although there seems to be a significant correlation between the features of a resilient system and an available one, one should bear in mind that they target different aspects of a trustworthy system. The similarity arises from the fact that both concepts are required for reliable systems. However, resilience emphasizes the system’s ability to recover from unexpected failures, while availability focuses on ensuring the system is accessible when needed. Thus, common features like redundancy, maintainability, and robustness are fundamental for achieving both resilient and available systems (Qadir & Quadri, 2016).

According to the research done by Mesbahi et al. (2018), disk failures are introduced as the major source of hardware failures in data centers. Considering the harsh working environment of autonomous ships, the risk of this issue increases compared to conventional data centers. Also, the risk of disk failure can have a more significant adverse impact on these types of ships due to their critical dependence on data for navigation, decision-making, and control systems; thus, any failures can result in a challenge in navigation safety. Also, as mentioned, physical access for ship maintenance is very challenging, emphasizing the importance of robust systems and advanced predictive maintenance techniques. In addition, in this paper, redundancy and reliability measures, such as real-time monitoring systems and sophisticated data backup and recovery systems, are suggested to ensure continuous operation in the event of failure.

Another solution for improving the availability in case of failures would be to benefit from cloud computing facilities. In (Endo et al., 2016), the authors systematically reviewed various high-availability solutions for cloud computing and discussed strategies for data replication across multiple servers to protect against data loss, monitoring, and recovery.

Security

Since AI systems usually deal with sensitive data, hardware security is of great importance. The systems should be protected from both physical tampering and cyber-attacks, such as side-channel attacks and hardware Trojans, to ensure data integrity and trust in decisions made by the application. Autonomous vessels are often targets of various cyber-attacks, and since the decisions are made based on the data and supporting algorithms, any cyber issues will risk the safety of ship navigation.

One acceptable practice to prevent these issues is implementing secure design in manufacturing hardware components, such as using chips resistant to side-channel attacks. Using trusted platform modules to store cryptographic keys securely and applying encryption are among the mentioned methods. However, some other solutions have been suggested by researchers to improve systems' security by using AI and ML approaches. For instance, Sayadi et al., (2022), focused on AI/ML techniques to improve hardware and architecture security as they can help identify potential vulnerabilities proactively. They introduced ML algorithms for hardware-assisted intrusion detection that identify and respond to security threats by monitoring and analyzing network traffic and system activities through hardware components. Also, deep learning methods for analyzing indirect, physical outputs, such as power consumption or electromagnetic emissions, can detect side-channel attacks during cryptographic operations.

Tan & Karri (2020) discussed in their paper how ML-based methods can have the potential to detect hardware Trojans in the design flow of hardware systems. Hardware Trojans are malicious functionalities inserted in the hardware that activate during the system runtime and can cause system disruption or damage by manifesting small changes in the system's design characteristics. It is proposed that ML methods are capable of detecting anomalies caused by Trojans by analyzing complex data patterns.

Moreover, among the methods discussed in (Tehranipoor, 2021) for enhancing hardware security, methods like blockchain for supply chain assurance and post-quantum hardware security can be seen. Blockchain is essential for preventing unauthorized components from entering the supply chain. In addition, post-quantum hardware security methods address the need for more advanced techniques since quantum computers would have the ability to break encryption methods in the future.

As discussed in (Hossain Faruk et al., 2022), while the advent of quantum computers is known to be a threat to the current cryptographic keys used for digital security, it can also provide new opportunities for enhancing cybersecurity through quantum key distribution and other quantum-resistant cryptographic methods such as lattice-based, code-based, hash-based, and multivariant-based cryptographic methods. These methods can be implemented on the hardware modules of autonomous ships to secure communications and data storage against post-quantum attacks. One can refer to (Mavroeidis et al., 2018) and other similar studies for further details.

Privacy

This feature significantly contributes to the trustworthiness of AI applications, and with the increasing incorporation of AI into people's daily lives, protecting user privacy is of utmost importance. To this end, data is usually encrypted before being stored or transmitted, reducing the risk of unauthorized access. It is believed that hardware encryption can be more secure compared to software encryption (Çetintav & Taşkın, 2023). Also, the data processing phase must be secured by techniques such as homomorphic encryption that allows the data to be processed while encrypted. Another recommendation would be to minimize the data exposure by analyzing them locally rather than in data centers to reduce the risk of data misuse. This method seems to be unfeasible for remotely controlled ships since the data are transmitted off the ship to the control centers for human oversights.

Following related privacy protocols such as the one suggested in ("Ethics Guidelines for Trustworthy AI. High-Level Expert Group on Artificial Intelligence," 2019) ensures a more trustworthy hardware system. According to this document, privacy in the field of AI systems is a crucial aspect that proper protocols must be considered to address the principle of prevention of harm. This principle emphasizes the importance of data governance which includes both data quality and integrity, access protocols, and data processing in a way that protects privacy. The quality and integrity of data used in AI applications are of paramount importance especially in cases with self-learning algorithms since feeding erroneous data may result in faulty decisions and jeopardize safety. Also, to protect against any sort of attacks that can affect the data integrity, a data access protocol is suggested to be developed. Based on this protocol, it is determined under which circumstances who can have access to the data and only qualified and responsible personnel are allowed to access it.

Safety

Discussing hardware system safety requires several specific considerations and addressing challenges. There must be clear, strict safety measures for the systems used in autonomous ships since they are highly dependent on AI to make decisions during operation. Some of these measures and the ways to improve them are already discussed in the previous sections. For example, the environmental robustness of the hardware systems, cybersecurity, reliable communication systems, and monitoring and maintenance are a few of the measures that have been touched upon.

In addressing the safety of MASS, several critical factors must be integrated into their design and operation. Firstly, fail-safe mechanisms are essential; in the event of system failure, the ship should automatically reduce speed or anchor to maintain safety. Additionally, considering the extended duration of voyages, the implementation of efficient power management becomes crucial. This can be achieved by utilizing energy-efficient components that guarantee consistent, uninterrupted operations. Furthermore, the necessity for real-time data processing demands that the hardware is not only capable of high-speed processing but also adept in complex data analysis. For ships controlled remotely, the integration of robust human-machine interfaces is vital to enable operators to seamlessly take control when necessary. Lastly, despite the absence of a crew, it is imperative that MASS are equipped with emergency response hardware, including firefighting systems and bilge pumps, to handle unforeseen incidents effectively. In conclusion, safety in autonomous ships is a multidimensional concept that must address various challenges to ensure the safe operation of the ships.

DISCUSSION

The study of trustworthiness in MASS from a hardware perspective along with the software aspect, is essential to ensure a safe transition from conventional ships to this type. System resilience, availability, security, privacy, and safety were identified as the key performance indicators (KPIs) of the hardware trustworthiness framework. Based on what was discussed in the previous sections, an interconnected relationship between these indicators is highlighted. Resilience in an AI system is the system’s ability to recover from failures. Resilience is closely connected to the concept of reliability, which can be defined as the system’s ability to perform its intended function consistently. Although availability is an independent concept, again, there is a close correlation between that and reliability. A system that is available and always ready to use can fulfill its reliability goal. Moreover, there are mutualities between security and privacy concepts. A secure application defends the system from unauthorized access while protecting sensitive data from being leaked. Thus, effective security measures are essential to maintain privacy.

In conclusion, enhancing the system's safety is inherently linked to the improvement of key features previously discussed. A resilient system adeptly manages unexpected conditions, thereby diminishing the likelihood of unsafe outcomes. Ensuring system availability safeguards critical safety measures, while reliability minimizes the occurrence of errors that could cause unsafe conditions. Moreover, the security of an AI system is crucial in preventing compromises that could lead to safety risks. Additionally, implementing robust privacy protection measures is essential to prevent data misuse, further contributing to the system's overall safety. Fig. 5 shows the inter-relationship between KPIs of hardware trustworthiness evaluation framework in summary.

Although key concepts for a trustworthiness evaluation of autonomous systems are identified, challenges remain in ensuring the smooth integration of these robust hardware systems with advanced AI algorithms. The balance between hardware resilience and software adaptability presents a unique challenge, especially in real-time decision-making scenarios encountered in maritime environments. Furthermore, this study provides opportunities for future research, especially in developing hardware and systems that can adapt and evolve in response to maritime threats and challenges in certain circumstances. As MASS continues to advance, the hardware behind these systems must meet current standards and be adaptable for future technological advancements.

CONCLUSION

Implementing AI-based digital navigators on autonomous ships introduces challenges to the shipping industry, highlighting the need for implementing a trustworthiness evaluation framework. This evaluation must cover both levels since digital navigators are developed by synthesizing software algorithms with respective hardware systems. This study serves as a step towards understanding and enhancing the hardware aspect of MASS, paving the way for safer and more reliable autonomous maritime operations.

This paper first discusses the steps for developing digital navigators and the systems used for their implementation. Then, KPIs for the hardware trustworthiness evaluation are identified as resilience, availability, reliability, security, privacy, and safety. Each of these concepts is the topic of each section in the paper, where a holistic overview is given. Then, based on the literature, some methods and techniques for enhancing these features applied in other industries are extracted and presented.

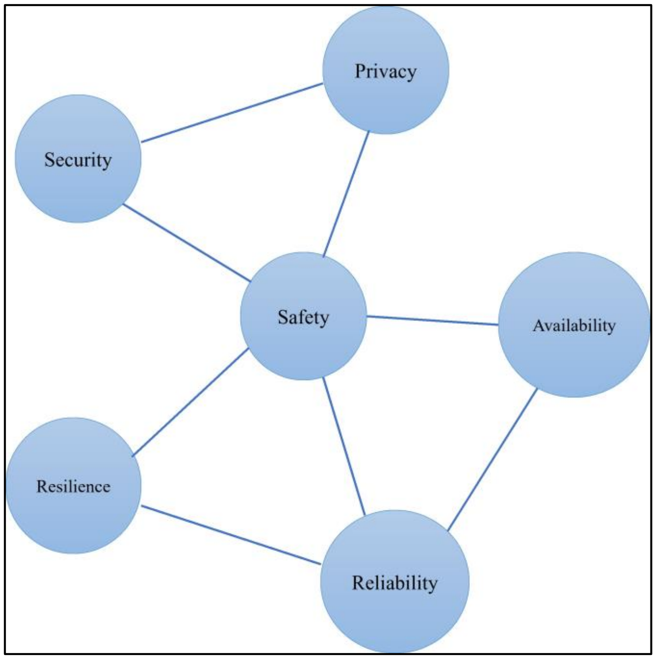

Fig 5. The inter-relationship between KPIs of hardware trustworthiness study framework

Autonomous vessels can have various possible applications, and as discussed one of the merits of these ship types is to reduce GH emissions. Since autonomous vessels may not have human life-supporting systems, the vessel hull and superstructure can be modified to improve ship energy efficiency, further. Moreover, such vessels can be facilitated with wave and wind energy harvesting devices, that can harvest renewal energy through their voyages. One should note that such vessels should navigate in moderate or rough weather conditions to harvest such renewable energy and that may not be idealistic conditions for humans. Since humans are not onboard, the respective energy efficiency can be maximized for such vessels to achieve the IMO emission reduction target for net zero-emission vessels by 2050. Furthermore, autonomous vessels can be used for various dangerous tasks that may have some risk to human operators, specifically in hard weather conditions.

ACKNOWLEDGMENTS

The work is a part of the Center of Excellence in Maritime Simulator Training and Assessment (COAST) originates from MARKOM and was established as the Center for Outstanding Education (SFU) in 2020.

This work is also supported by the MARKOM II project under the project title "Onshore Operation Center for Remotely Controlled Vessels (OOC 2023)" under contract number PMK-2022-10014.

REFERENCES

Agarwal, A., & Agarwal, H. (2022). A Seven-Layer Model for Standardising AI Fairness Assessment. *ArXiv Preprint ArXiv:2212.11207*.

Apostol-mates, R., & Barbu, A. (2016). HUMAN ERROR-THE MAIN FACTOR IN MARINE ACCIDENTS. *Scientific Bulletin of*

Naval Academy, *19*(2), 451–454. https://doi.org/10.21279/1454-864X-16-I2-068

Barredo, A., Díaz-rodríguez, N., Del, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts , taxonomies , opportunities and challenges toward responsible AI. *Information Fusion*, *58*(October 2019), 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Bolchini, C., Cassano, L., & Miele, A. (2023). *Resilience of Deep Learning applications: a systematic survey of analysis and hardening techniques*. *Ml*, 1–32. https://doi.org/https://doi.org/10.48550/arXiv.2309.16733

Çetintav, I., & Taşkın, D. (2023). Performance analysis of hardware and software based AES encryption on internet of things SoC. *AIP Conference Proceedings*, *2849*(1).

Cheng, L., Kush, R. V., & Liu, H. (2021). Socially Responsible AI Algorithms : Issues , Purposes , and Challenges. *Journal of Artificial Intelligence Research*, *71*, 1137–1181.

Dave, A., Banerjee, N., & Patel, C. (2021). CARE: Lightweight Attack Resilient Secure Boot Architecture with Onboard Recovery for RISC-V based SOC. *2021 22nd International Symposium on Quality Electronic Design (ISQED)*, *2021-April*, 516–521. https://doi.org/10.1109/ISQED51717.2021.9424322

Endo, P. T., Rodrigues, M., Gonçalves, G. E., Kelner, J., Sadok, D. H., & Curescu, C. (2016). High availability in clouds: systematic review and research challenges. *Journal of Cloud Computing*, *5*(1), 16. https://doi.org/10.1186/s13677-016-0066-8

Eriksen, S., & Lützen, M. (2022). The impact of redundancy on reliability in machinery systems on unmanned ships. *WMU Journal of Maritime Affairs*, *21*(2), 161–177. https://doi.org/10.1007/s13437-021-00259-7

Ethics guidelines for trustworthy AI. High-Level Expert Group on Artificial Intelligence. (2019). *European Commission*, 1–39. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Hershey, R. H. (1956). Reliability Through Redundancy. *IRE Transactions on Aeronautical and Navigational Electronics*, *ANE-3*(1), 16–20. https://doi.org/10.1109/TANE3.1956.4201435

Hossain Faruk, M. J., Tahora, S., Tasnim, M., Shahriar, H., & Sakib, N. (2022). A Review of Quantum Cybersecurity: Threats, Risks and Opportunities. *2022 1st International Conference on AI in Cybersecurity (ICAIC)*, 1–8. https://doi.org/10.1109/ICAIC53980.2022.9896970

Huang, Y., Chen, L., Negenborn, R. R., & van Gelder, P. H. A. J. M. (2020). A ship collision avoidance system for human-machine cooperation during collision avoidance. *Ocean Engineering*, *217*, 107913. https://doi.org/10.1016/j.oceaneng.2020.107913

IMO. (2018). Maritime Safety Committee for the regulatory scoping exercise for the use of maritime autonomous surface ships (MASS). *Maritime Safety Committee 100th Session, MSC 100/ WP.8*.

Junior, F. M. R., & Kamienski, C. A. (2021). A Survey on Trustworthiness for the Internet of Things. *IEEE Access*, *9*, 42493–42514. https://doi.org/10.1109/ACCESS.2021.3066457

Kim, T. eun, Perera, L. P., Sollid, M. P., Batalden, B. M., & Sydnes, A. K. (2022). Safety challenges related to autonomous ships in mixed navigational environments. *WMU Journal of Maritime Affairs*, *21*(2), 141–159. https://doi.org/10.1007/s13437-022-00277-z

Lipton, Z. C. (2018). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, *16*(3), 31–57. https://doi.org/10.1145/3236386.3241340

Mavroeidis, V., Vishi, K., D., M., & Jøsang, A. (2018). The Impact of Quantum Computing on Present Cryptography. *International Journal of Advanced Computer Science and Applications*, *9*(3), 405–414. https://doi.org/10.14569/IJACSA.2018.090354

Mesbahi, M. R., Rahmani, A. M., & Hosseinzadeh, M. (2018). Reliability and high availability in cloud computing environments: a reference roadmap. *Human-Centric Computing and Information Sciences*, *8*(1), 20. https://doi.org/10.1186/s13673-018-0143-8

Montavon, G., Lapuschkin, S., Binder, A., Samek, W., & Müller, K. (2017). Explaining nonlinear classification decisions with deep Taylor decomposition. *Pattern Recognition*, *65*(August 2016), 211–222. https://doi.org/10.1016/j.patcog.2016.11.008

Murray, B., & Perera, L. P. (2021). An AIS-based deep learning framework for regional ship behavior prediction. *Reliability Engineering & System Safety*, *215*(May), 107819. https://doi.org/10.1016/j.ress.2021.107819

Namazi, H., & Perera, L. P. (2023). Trustworthiness Evaluation Framework for Digital Ship Navigators in Bridge Simulator Environments. *Proceedings of the International Conference on Offshore Mechanics and Arctic Engineering - OMAE*, *5*(June). https://doi.org/10.1115/OMAE2023-104863

Perera, L. P. (2020). Deep Learning Toward Autonomous Ship Navigation and Possible COLREGs Failures. *Journal of Offshore Mechanics and Arctic Engineering*, *142*(3), 1–39. https://doi.org/10.1115/1.4045372

Qadir, S., & Quadri, S. M. K. (2016). Information Availability: An Insight into the Most Important Attribute of Information Security. *Journal of Information Security*, *7*(3).

Sayadi, H., Aliasgari, M., Aydin, F., Potluri, S., Aysu, A., Edmonds, J., & Tehranipoor, S. (2022). Towards AI-Enabled Hardware Security: Challenges and Opportunities. *2022 IEEE 28th International Symposium on On-Line Testing and Robust System Design (IOLTS)*, 1–10. https://doi.org/10.1109/IOLTS56730.2022.9897507

Taghavi, M., & Perera, L. P. (2022). Data Driven Digital Twin Applications Towards Green Ship Operations. *Volume 5A: Ocean Engineering*, 1–10. https://doi.org/10.1115/OMAE2022-78775

Taghavi, M., & Perera, L. P. (2023). Multiple Model Adaptive Estimation Coupled With Nonlinear Function Approximation and Gaussian Mixture Models for Predicting Fuel Consumption in Marine Engines. *Volume 5: Ocean Engineering*, *Ml*, 1–9. https://doi.org/10.1115/OMAE2023-103249

Tan, B., & Karri, R. (2020). Challenges and New Directions for AI and Hardware Security. *2020 IEEE 63rd International Midwest*

Symposium on Circuits and Systems (MWSCAS), *2020-Augus*, 277–280. https://doi.org/10.1109/MWSCAS48704.2020.9184612

Tehranipoor, M. (2021). Emerging Topics in Hardware Security. In M. Tehranipoor (Ed.), *Springer*. Springer International Publishing. https://doi.org/10.1007/978-3-030-64448-2

Vartdal, B. J., Skjong, R., & St.Clair, A. L. (2018). *Remote-controlled and autonomous ships in the maritime industry*.

Wan, Z., Swaminathan, K., Chen, P.-Y., Chandramoorthy, N., & Raychowdhury, A. (2022). Analyzing and Improving Resilience and Robustness of Autonomous Systems. *Proceedings of the 41st IEEE/ACM International Conference on Computer-Aided Design*, 1–9. https://doi.org/10.1145/3508352.3561111

Zeiler, M. D., & Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Computer Vision -- ECCV 2014* (pp. 818–833). Springer International Publishing.