IALA GUIDELINE

G1178
AN INTRODUCTION TO ARTIFICIAL INTELLIGENCE (AI) FROM AN IALA PERSPECTIVE

Edition 1.0

December 2022

urn:mrn:iala:pub:g1178:ed1.0

10, rue des Gaudines – 78100 Saint Germain en Laye, France
Tél. +33 (0)1 34 51 70 01 – contact@iala-aism.org

www.iala-aism.org

International Association of Marine Aids to Navigation and Lighthouse Authorities
Association Internationale de Signalisation Maritime

DOCUMENT REVISION

Revisions to this document are to be noted in the table prior to the issue of a revised document.

Date	Details	Approval
December 2022	First issue	Council 76

CONTENTS

List of Figures and Tables

1. BACKGROUND

An artificial Intelligence (AI) system is a machine-based system that can, for a given set of defined objectives, make predictions, recommendations, or decisions. AI systems offer functionality needed to operate with varying levels of autonomy [1].

Deep learning, machine learning and AI are all related to each other. The learning methods make use of large amounts of data. This results in a performance that often cannot be achieved using classical discrete algorithms. The amount of data needed leads to questions related to data privacy.

There are concerns that need to be considered by regulators, providers, and users of maritime centric artificial intelligence systems. These are often addressed by policy or guidelines that are organization centric. This Guideline is a living document and seeks to provide guidance in consideration of AI within the IALA domain.

1.1. OBJECTIVE

Understand the advantages and risks of AI within the IALA domain and how to manage this risk now, and provide guidance going forwards, recognizing the rapid growth of AI and its capabilities.

1.2. SCOPE

The following topics are in scope of this document:

- applications used in maritime environment for AtoN and VTS (within the IALA mandate); and
- AI with Machine Learning (ML) and deep learning.

The evaluation and recommendation of commercial AI and ML solutions are out of scope.

2. OVERVIEW

The relationship between deep learning, neural networks, ML and AI is often diagrammatically explained as provided in Figure 1. This is based on the Organisation for Economic Co-operarion and Development (OECD) AI Principles [2] and often termed “AI subsets”.

Often significant volumes of data are required to train the AI models to enable the required outcomes to be achieved for example, many pictures of person's faces are required to enable an AI model that can deal accurately with facial recognition.

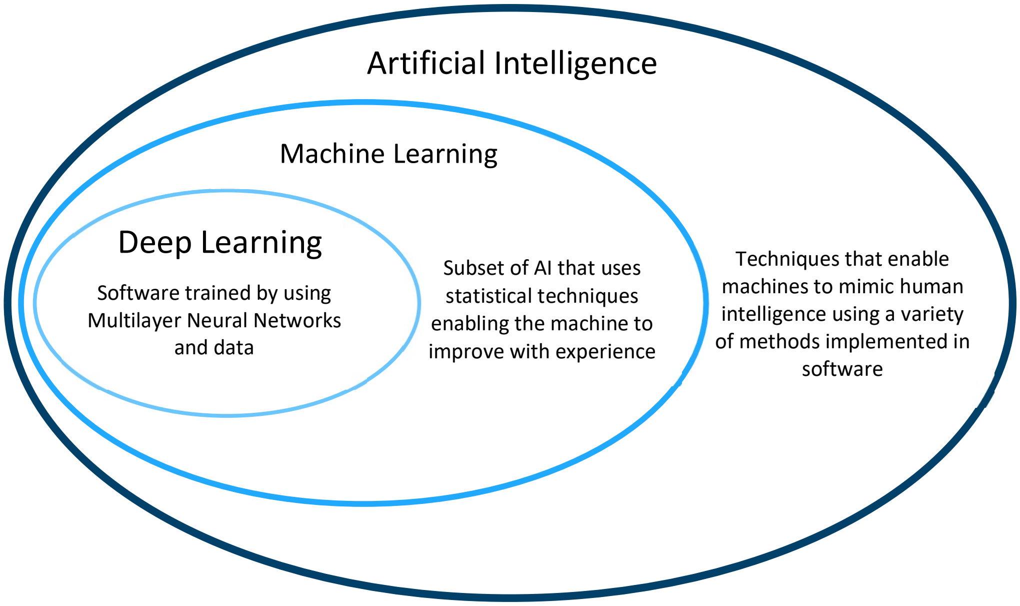

Figure 1 Overview of AI and subsets (adapted from OECD [2])

The concerns that surround AI systems are:

1. Bias (e.g., commercial, cultural and gender);
2. Accuracy (e.g., AI versus deterministic systems);
3. Transparency (e.g., when investigating an incident, how is IALA to deal with AI decision support tools);
4. State of the AI system when a decision is made;
5. Conflict between different AI systems in the same domain;
6. The patenting of AI systems;
7. The commercial value of working and tested AI systems; and
8. Data privacy issues.

2.1. BIAS

Bias includes potential cultural, gender, race, and commercial biases. Aspects of bias will reflect the data used for training.

For example, these biases may become apparent when AI is used to detect which VTS operator is at a specified operating position, the operator's attentiveness, and activity over the period of a shift.

2.2. ACCURACY

Accuracy of AI is identified through a confidence score. A confidence score is a number between 0 and 1 that represents the likelihood that the output of a machine learning model is correct and will satisfy a user's request, where 1 represent a 100% accuracy.

As an example, using Microsoft's breakdown of confidence score's meaning for an interactive voice response system can be categorized as follows:

- Over 0.7: the prediction is a strong candidate for answering the user query.
- Between 0.3 and 0.7: the prediction can partially answer the request.
- Below 0.3: the prediction is probably not a good choice.

In practical applications, confidence values close to 1 are required in safety critical systems. The new methods used in AI and ML require new approaches to test and evaluate the confidence in, and accuracy of, AI based systems. This means that new approaches are required for audit and certification of AI systems.

2.3. TRANSPARENCY

IALA has traditionally used systems that are deterministic. This means that the systems are rules based and, for the same inputs, the same output is guaranteed. Users should be sure of the same result when AI systems are used.

Transparency is required when decisions are made in an AI system where the user has no insight as to how the decision was made.

AI methods are combined statistical methods. The decisions made by AI are usually not explainable. The statistical approach implemented in the AI, and the data used for training in the AI system, should be declared and the approach should be able to be explained.

2.4. STATE OF AI WHEN A DECISION IS MADE

The accuracy of an AI system relies on the implementation of a statistical approach. In addition, the accuracy is based on the quantity and quality of the data used in training and any new data that has been processed. The state of the AI when a decision is made should be auditable (see APPENDIX 1).

2.5. CONFLICT OF SYSTEMS IN THE SAME DOMAIN

Each AI system relies on the implementation of different AI statistical approaches. They can also use different training data, with access to different sensors and live data streams, working to deliver similar outputs.

For example, ship route optimization will use several different data sources, including port of departure, type of ship and weather forecasts along the route. This will allow an optimized route to be provided. Different route optimization companies use different AI approaches, which may provide different routes.

2.6. PATENTS

This refers to the need to declare the patent status of various systems used in the IALA domain, which should include declaration of AI systems used. This may include a subset of a patent, or different providers may refer to the same AI patent.

2.7. COMMERCIAL VALUE

This refers to the commercial value of AI systems. The value of some AI systems can be very high and the protection of these, as well as the associated training data, needs to be considered.

3. BENEFITS AND CHALLENGES OF AI WITHIN THE IALA CONTEXT

There are identified benefits and challenges with the use of AI. These have been adapted to the IALA context and include those identified in Table 1.

Table 1 Examples and Challenges of AI in the IALA Context

Examples	Challenges
– Ensuring VTS operator focus – Approach and departure management – AtoN system availability management for both maintenance and service availability – Use of AI in Radar target extraction – AI to detect ships using CCTV cameras – VTS situational awareness	– Applicable data – Transformation of processes – AI training processes – Gathering the data for the data model – Audit regime for AI

4. AUDIT REGIME FOR AI

As AI becomes integrated into systems and processes, it is important that they are subject to a comprehensive auditing regime. The primary aspects of the audit will include compliance (assessing the risk related to the use of the AI and compliance with standards) and technology (assessing the risk related to the AI itself, privacy and security of data). AI should be implemented in a manner that supports an audit regime.

A proposed approach for an auditing regime of AI is provided in APPENDIX 1. This will need to be adapted and expanded further within the IALA context as experience is gained in the implementation of AI in the IALA context.

5. CONCLUSION

As AI grows in usage in the maritime domain, IALA has a responsibility to consider how the use of this technology can assist and affect the IALA members. Some guiding principles include:

1. AI systems should make sure that AI-driven decisions are fair and free of any harmful bias and endeavour to develop AI in an ethical way so that it can be trusted. This should also ensure that outputs from these data-driven systems support effective decision making and do not guide users to make decisions that may affect any group or individual in an unfair way.
2. AI systems should promote transparency and accountability. AI systems should inform users when they communicate directly with AI-powered systems and/or are subject to outcomes in which AI system have played a role.
3. Designers and providers of AI systems should endeavour to respect the privacy and protect the security of all individuals served by the AI systems deployed.
4. AI systems should be designed and deployed in a manner that fosters diversity, accessibility, and inclusivity.
5. AI systems should be designed and deployed in a manner that supports investigations of incidents by using audit mechanisms during design, development, deployment, and operation.

6. DEFINITIONS

The definitions of terms used in this Guideline can be found in the *International Dictionary of Marine Aids to Navigation* (IALA dictionary)and were checked as correct at the timeof going to print. Where conflict arises, the IALA Dictionary should be considered as the authoritative source of definitions used in IALA documents.

Additional definitions specific to this document are:

Artificial intelligence	An artificial intelligence (AI) system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.
Machine learning	Machine learning (ML) is the use and development of computer systems that are able to learn and adapt without following explicit instructions by using algorithms and statistical models to analyse and draw inferences from patterns in data.
Deep learning	Deep learning is a type of machine learning based on artificial neural networks in which multiple layers of processing are used to extract progressively higher-level features from data.

7. ABBREVIATIONS

AI	Artificial intelligence
ML	Machine learning
OECD	Organization for Economic Co-operation and Development

8. REFERENCES

[1] https://www.oecd.ai/ai-principles.

[2] OECD (2021), Artificial Intelligence, Machine Learning and Big Data in Finance: Opportunities, Challenges, and Implications for Policy Makers, https://www.oecd.org/finance/artificial-intelligence-machine-learningbig-data-in-finance.htm.

9. FURTHER READING

[1] Chong, Leah. (2021) Human confidence in artificial intelligence and in themselves: The evolution and impact of confidence on adoption of AI advice. https://www.sciencedirect.com/science/article/pii/S0747563221003411.

[2] https://www.vodafone.com/about-vodafone/how-we-operate/public-policy/policy-positions/artificial-intelligence-framework.

[3] http://www.g7.utoronto.ca/summit/2018charlevoix/ai-commitment.html.

[4] https://medium.com/voice-tech-global/machine-learning-confidence-scores-all-you-need-to-know-as-a-conversation-designer-8babd39caae7.

[5] https://dataconomy.com/2022/04/is-artificial-intelligence-better-than-human-intelligence/.

APPENDIX 1 SAMPLE AI AUDIT FRAMEWORK

An initial internal audit framework can be framed as encompassing five distinct stages - Scoping, Mapping, Artefact Collection, Testing and Reflection (SMACTR) - all of which have their own set of documentation requirements and account for a different level of the analysis of a system[1].

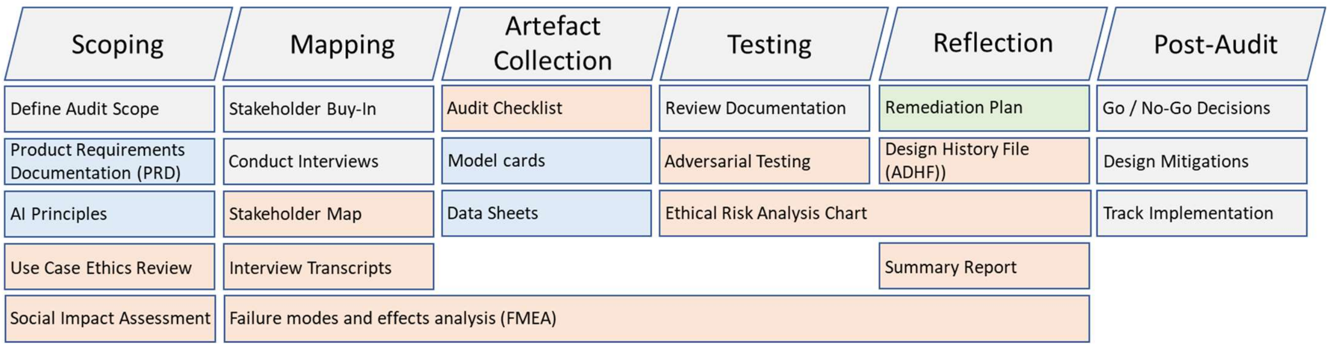

Figure 2 Sample AI Audit Model

1.1. SCOPING STAGE

This is the stage in which the risk analysis begins by mapping out intended use cases and identifying analogous deployments either within the organization or from competitors or adjacent industries. The goal is to anticipate areas to investigate as potential sources of harm and social impact. At this stage, interaction with the system should be minimal.

1.2. MAPPING STAGE

This is a review of what is already in place and the perspectives involved in the audited system. This is also the time to map internal stakeholders, identify key collaborators for the execution of the audit, and orchestrate the appropriate stakeholder buy-in required for execution.

1.3. ARTEFACT COLLECTION STAGE

This stage requires the identification and collection all the required documentation from the product development process, to prioritize opportunities for testing and can include other product development artifacts such as design documents and reviews, in addition to systems architecture diagrams and other implementation planning documents and retrospectives.

1.4. TESTING STAGE

This stage is when the auditors execute a series of tests to gauge the compliance of the system with the prioritized ethical values of the organization. Auditors engage with the system in various ways and produce a series of artifacts to demonstrate the performance of the analysed system at the time of the audit. Additionally, auditors review the documentation collected from the previous stage and begin to make assessments of the likelihood of system failures to comply with declared principles.

1.5. REFLECTION STAGE

This phase of the audit is the more reflective stage, when the results of the tests at the execution stage are analysed in juxtaposition with the ethical expectations clarified in the audit scoping. This phase will reflect on product decisions and design recommendations that could be made following the audit results.

---

1 https://doi.org/10.1145/3351095.3372873