|  |
| --- |
| IALA Guideline |

Artificial Intelligence Guideline

(Working title)

Task Proposal:

1. *Understand the AI terminology (operational perspective)*
2. *Understand the pitfalls (Operational perspective)*
3. *Understand how AI systems are audited*
4. *Align the AI terminology with the IALA terminology*
5. *Generate an IALA Guideline (similar to AI policies in industry)*
6. *Generate an IALA Recommendation*
7. *Generate an IALA centric AI audit that supports the IALA Guideline and Recommendation on AI*

Edition x.x

Date (of approval by Council)

urn:mrn:iala:pub:gnnnn

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

|  |  |  |
| --- | --- | --- |
| Date | Details | Approval |
| 28th October 2021 | First Draft |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

1. Background 6

1.1. Objective 6

1.2. Scope 6

2. Overview 6

2.1. Relationship 6

2.2. Bias 7

2.3. Accuracy 7

2.4. Black Box syndrome 7

2.5. State of AI 7

2.6. Transparency 7

2.7. Conflict of systems in the same domain 7

2.8. Patents 8

2.9. Commercial Value 8

2.10. other??? 8

3. Benefits and Challenges of AI within the IALA Context 8

4. Audit regime for AI 8

5. Conclusion 8

6. DEFINITIONS 9

7. abbreviations 9

8. references 10

9. Further reading 10

10. Index 10

11. APPENDIX 1 – AN AI Audit framework 11

11.1. Scoping stage: 11

11.2. Mapping stage: 11

11.3. Artefact collection stage: 11

11.4. Testing stage: 11

11.5. Reflection stage: 11

List of Figures

Figure 1 Artificial Intelligence 4

Figure 2 AI Audit Model 7

# Background

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[[1]](#footnote-1).

Deep Learning, Machine Learning and Artificial Intelligence are all related to each other. The learning methods used allow the processing of large amounts of data and results in a performance that cannot be achieved by humans when confronted by the same large volume of data.

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

## Objective

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

## Scope

Applications used in maritime environment and AtoN and VTS (within the IALA mandate)

Refer to Task Register document for further input.

[what is in scope / what is out of scope?]

# Overview

[introductory text regarding development ANN, and importance of AI, ML, Deep learning]

## Relationship

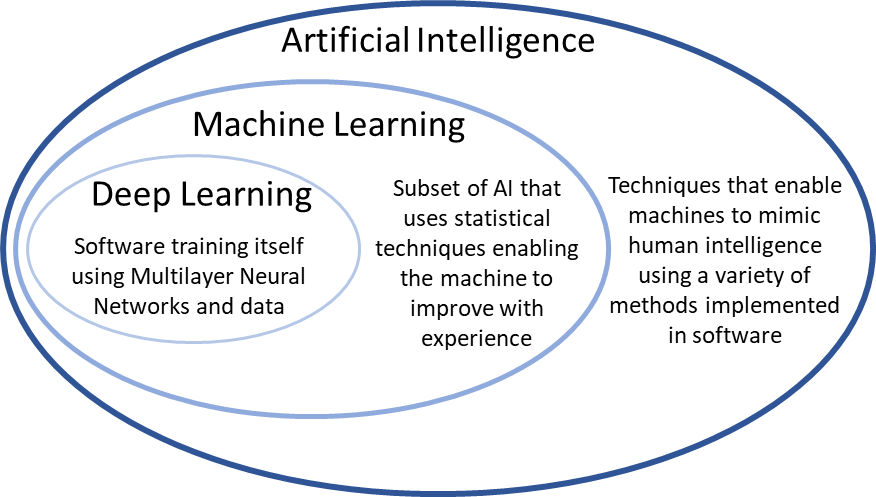
The relationship between Deep Learning, Neural Networks, Machine Learning (ML) and Artificial Intelligence (AI) is often diagrammatically explained as shown below[[2]](#footnote-2). Big Data and its role in AI?

Figure 1 Artificial Intelligence

The concerns that surround AI systems are:

1. Bias
2. Accuracy
3. Black Box Syndrome
4. State of the Artificial Intelligence system when a decision is made
5. Transparency
6. Conflict between different Artificial Intelligence systems in the same domain
7. The patenting of Artificial Intelligence systems
8. The commercial value of working and tested AI systems

## Bias

Take into account potential cultural, gender, race and commercial biases.

Note: Concerns about how AI works and its use case in the maritime domain.

## Accuracy

Deal with false positives and negatives.

## Black Box syndrome

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

What is the black box syndrome?

When decisions are made in an AIS system where the user has no insight as to how the decision was made.

## State of AI

The accuracy of an AI system, besides using the correct algorithm, is based on the quantity and quality of the data used in training and what new data has been processed. The outputs of the AI system with its inputs should be able to be explained.

Is addition explanation required?

## Transparency

The algorithms and data used for training should be declared and the algorithm should be able to be explained.

Is this the same topic as the Black Box Syndrome?

## Conflict of systems in the same domain

Each AI system has a different AI algorithm and can use different training data and has access to different sensors with live data streams but should deliver the same outputs.

Expand using a use case as explanation.

## Patents

The need to declare the patent status of various systems used in the IALA domain should include AI systems.

## Commercial Value

The commercial value of some AI systems can be very high and the protection of these and the associated training data needs to be considered

## other???

# Benefits and Challenges of AI within the IALA Context

Identify key benefits / challenges – perhaps do a SWOT style approach for the use of AI within iALA?

Link to compelling need in the task proposal:

AI is being made part of Decision Support Tools and work is ongoing that includes:

1. Ensuring VTS operator focus
2. Approach and departure management
3. AtoN system availability management

# Audit regime for AI

Tasks definition components include:

1. Understand the pitfalls (Operational perspective)
2. Understand how AI systems are audited

See Appendix A

# Conclusion

As AI grows in usage in the maritime domain IALA has a responsibility to consider how the use of this technology affects the IALA members. The guiding principles are as follows:

1. AI systems should make sure AI-driven decisions that are fair and free of any harmful bias.
2. AI systems should promote Transparency and Accountability and inform users when they communicate directly with AI-powered systems and / or are subject to outcomes in which AIS system have played a role.
3. Endeavour to develop AI in an ethical way so that it can be trusted ensuring that outputs from these data-driven systems do not inadvertently guide users to make decisions that may affect any group or individual in an unfair way.
4. AI systems will strive to ensure that there is effective oversight and a ‘human-in-control’ approach to the use of AI.
5. Designers and providers of AIS systems should endeavour to respect the privacy and protect the security of all individuals served by the AI systems deployed.
6. AI systems should be designed and deployed in a manner that fosters diversity, accessibility, and inclusivity.
7. 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.

# DEFINITIONS

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 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.

Other definitions of terms used in this Guideline can be found in the *International Dictionary of Marine Aids to Navigation* (IALA Dictionary) at <http://www.iala-aism.org/wiki/dictionary> and were checked as correct at the time of going to print. Where conflict arises, the IALA Dictionary should be considered as the authoritative source of definitions used in IALA documents.

# abbreviations

AI Artificial Intelligence

ML Machine Learning

# references

# Further reading

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

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

# Index

**No index entries found.**

# APPENDIX 1 – AN AI Audit framework

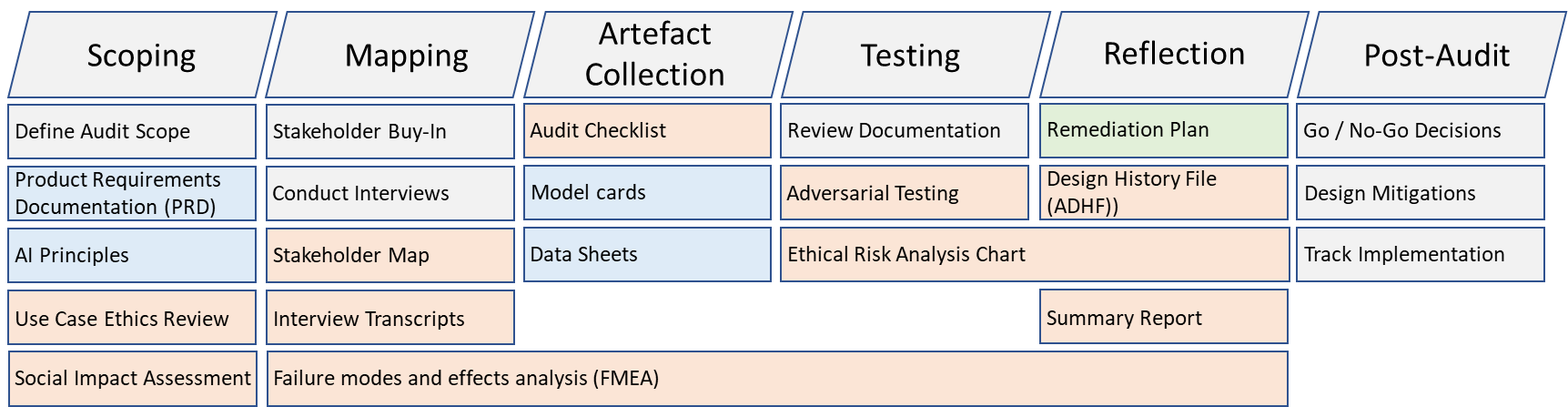
An initial internal audit framework can be framed as encompassing five distinct stages - Scoping, Mapping, Artefact Collection, Testing and Reflection (SMACTR) - all of which have their own set of documentation requirements and account for a different level of the analysis of a system[[3]](#footnote-3).

Figure 2 AI Audit Model

## 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.

## 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.

## Artefact collection stage:

This stage requires the identification and collection all the required documentation from the product development process, to prioritise 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.

## Testing stage:

This stage is when the auditors execute a series of tests to gauge the compliance of the system with the prioritised ethical values of the organisation. 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.

## 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://www.oecd.ai/ai-principles [↑](#footnote-ref-1)
2. https://www.ibm.com/blogs/systems/ai-machine-learning-and-deep-learning-whats-the-difference/ [↑](#footnote-ref-2)
3. https://doi.org/10.1145/3351095.3372873 [↑](#footnote-ref-3)