Input paper: [[1]](#footnote-1) ENAV28-5.1.2.1

Input paper for the following Committee(s): check as appropriate Purpose of paper:

**□** ARM **□** ENG **□** PAP **□** Input

X ENAV **□** VTS **□** Information

Agenda item [[2]](#footnote-2) 5.n

Technical Domain / Task Number 2 …………………………………

Author(s) / Submitter(s) …………………………………

Artificial Intelligence, Machine Learning and Deep Learning within the domain of IALA

# Summary

Artificial Intelligence or AI is being promoted as having a similar impact on the world as did the industrial revolution (so called Transformative Artificial Intelligence (TAI)). As this the use of Artificial Intelligence expands, we often hear the words, ‘data is the new oil’. But what is the expected impact on maritime Aids to Navigation?

An 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. https://www.oecd.ai/ai-principles

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.

The learning processes leads to the Artificial Intelligence ‘Black Box’ concern where the human users of these systems no longer understand the algorithm leading to the outcomes presented to the human and how these outcomes were arrived at.

The human can learn to trust the Artificial Intelligence system especially when its output appears to have been correct over many days, weeks, and months. But there are concerns that need to be considered by regulators, providers and users of maritime centric Artificial Intelligence system and include:

1. Bias – the bias such as race, gender and commercial interest is often dependent on the training model and data.
2. Accuracy – it is possible to have false positives and negatives.
3. State of the Artificial Intelligence system when a decision is made – the learning model may be dynamic and thus what is today may be different to what was at any other time.
4. Conflict between different Artificial Intelligence systems in the same domain – one Artificial Intelligence system gives one output and another, in the same domain with either different input or training data, gives a different output.
5. The patenting of Artificial Intelligence systems – the ownership the outcome of any one Artificial Intelligence system that is the outcome of a specific set of training data could be different to the same system using different training data.
6. The potential value of working and tested AI systems in any domain may lead to a ‘land grab’ by commercial companies, governments, and other interested parties.

Several potential AtoN use cases can be described where Artificial Intelligence initially informs the human and then augments the human and then replaces the human. Along the way we could see the development of products and systems that seem not to be viable or possible today.

# Purpose of the document

Diagram, circle

Description automatically generatedThe relationship between Deep Learning, Neural Networks, Machine Learning (ML) and Artificial Intelligence (AI) is often diagrammatically explained as shown below[[3]](#footnote-3).

For the purposes of this paper, an understanding of the technicalities of the methods employed in AI systems[[4]](#footnote-4) is not important or required.

# Discussions

## Bias

What is important is that where 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. This may not be true for AI in two respects:

1. The AI outcome is dependent on the data that it was trained on. This has meant that facial recognition systems trained using as select portion of the global population performs badly or with bias for other parts of the global population[[5]](#footnote-5). This has led to racial and gender bias being detected in AI based facial recognition systems.
2. The AI systems are often designed to continually learn and thus improve. This means that the AI system that was in place yesterday may be different to the system in place today in some small way or another. The changes over time will be greater as the amount of data used for the AI learning increases.

The growing issue is that AI is viewed by most users as a ‘Black Box’[[6]](#footnote-6) that cannot be easily explained to a typical third-party or layman user and so trust in the AI system is primarily determined by measuring the output and use over time. Where the output measurement, an AI Audit[[7]](#footnote-7), does not take account of any bias (gender, race, commercial, cultural, national, etc.), the bias may not be easily detected.

## Accuracy

The continuous learning capability of AI may make incident investigation difficult given that the AI system is learning and thus potentially changing from day to day, week to week and month to month[[8]](#footnote-8). The traditional data recording of the sensor data used for incident investigations may now not be sufficient to provide the required insights and learnings that were possible in the past because only the input and outputs are able to be recorded but how the outputs were arrived at is now lost because the state of the AI system is potentially difficult to record and then replicate[[9]](#footnote-9).

A picture containing text, light

Description automatically generatedAs the AI system is used and fed with additional data, the AI system is learning and, especially in the early stages of learning and deployment, adjusting its responses based on the training data[[10]](#footnote-10). This means that AI system’s performance at one point in time may be different to another point in time. Based on the image below, once the algorithm is set, these changes are primarily driven by the amount of data being used in the learning process. The more data, the higher the performance as shown in the diagram below.

The accuracy of an AI system, besides using the correct algorithm, is based on the quantity and quality of the data used in training. If the training data had issues (known or unknown), then that AI system will be trained using the faulty data. It is also true that the data being used at any one time by the AI system is often assumed to be trusted data[[11]](#footnote-11). Not all data can be trusted for reasons that include:

1. Faulty sensors feeding the AI system.
2. Data that has been processed or stored incorrectly or deliberately altered before being used for training.
3. The AI system being the subject of a Cyber Attack[[12]](#footnote-12) that changes the data being fed to the AI system to change the outcome sufficiently to achieve the attacker’s goal which could be difficult to ascertain in the short term and thus mitigate.

Some AI systems are challenged to achieve 100% accuracy, and this includes complex image recognition (examples are facial recognition and breast cancer recognition where 80% to 90% is considered good performance[[13]](#footnote-13)). Where deterministic algorithms already exist and have known 100% accuracy (examples are ship in an area / not in an area, ship exceeding a speed limit / ship not exceeding a speed limit), the role and / or claimed advantages of AI may be questioned.

## Human in the loop

When considering shore-based systems and the ‘human in the loop’ [[14]](#footnote-14) [[15]](#footnote-15)concept, the first implementations are going to be as indicated as Case A where the AI supports the human and fulfils the role of a Decision Support Tool (DST). As time moves on and the capability and trust in AI increases, as indicated in Case B, the human is going to monitor and manage the AI system.

The last and final step is as indicated in Case C where the human outputs become part of the inputs to the AI system and the AI system is seen to be the expert in the narrowly defined functionality of the considered shore-based system.

A picture containing diagram

Description automatically generated

It is probable that humans, in some capacity or another, may always be in the loop[[16]](#footnote-16).

## Conflicting outputs

In a hypothetical scenario as shown in the following image, there could be an AI system on the Maritime Autonomous Surface Ship (MASS), another that supports or provides the pilotage and another that supports or provides information to the Vessel Traffic Services (VTS).

A screenshot of a computer

Description automatically generated with low confidenceEach system has a different AI algorithm, used different training data, and has access to different sensors with live data streams. In this hypothetical example there are now three different AI systems giving three different results for the same situation.

This will need to be resolved by some mechanism that could include:

1. Allocating priority in various voyage phases to a particular AI system.
2. Allocating priority in various voyage phases to the AI system that has been audited and certified by some mechanism.
3. Allocating responsibility for the vessel depending on the phase of the voyage and there is no priority system.

This thinking can be expanded further. If the same AI algorithm is trained with three different data sets and one data set proves to be more accurate, then that AI system has a performance and thus commercial advantage. This also means that, although the same AI algorithm is used, the training data (historical and real time) can produce different results.

## AI Patent considerations

The above could also mean that a trained AI system using a particular data set can be considered to be Intellectual Property (IP) that can be patented and falls within the Intellectual Property Rights (IPR) considerations of IALA and other similar maritime centric organisations. The same could probably be said of the data set used for the training although this may have been derived from a publicly available data set (i.e., the global Satellite AIS data set).

There have already been one or more patent applications where AI is the named inventor[[17]](#footnote-17) (most of these appear to have failed thus far).

# Annexes

## Annex 1 – AI Policy – Charlevoix Common Vision for the Future of Artificial Intelligence

Leaders of the G7 – Canada, France, Germany, Italy, Japan, the United Kingdom, and the United States – met in Charlevoix, Canada in June 2018 and committed to the “Charlevoix Common Vision for the Future of Artificial Intelligence.”[[18]](#footnote-18) This shared vision highlights the potential for economic growth alongside pressing societal challenges; it states, “AI that fosters economic growth, societal trust, gender equality and inclusion depends on a predictable and stable policy environment that promotes innovation.”

The vision includes 12 commitments made by the G7 leaders. These commitments are available in full below:

1. Endeavour to promote human-centric AI and commercial adoption of AI, and continue to advance appropriate technical, ethical and technologically neutral approaches by: safeguarding privacy including through the development of appropriate legal regimes; investing in cybersecurity, the appropriate enforcement of applicable privacy legislation and communication of enforcement decisions; informing individuals about existing national bodies of law, including in relation to how their personal data may be used by AI systems; promoting research and development by industry in safety, assurance, data quality, and data security; and exploring the use of other transformative technologies to protect personal privacy and transparency.
2. Promote investment in research and development in AI that generates public trust in new technologies and encourage industry to invest in developing and deploying AI that supports economic growth and women’s economic empowerment while addressing issues related to accountability, assurance, liability, security, safety, gender and other biases and potential misuse.
3. Support lifelong learning, education, training and reskilling, and exchange information on workforce development for AI skills, including apprenticeships, computer science and STEM (science, technology, engineering, and mathematics) education, especially for women, girls and those at risk of being left behind.
4. Support and involve women, underrepresented populations and marginalized individuals as creators, stakeholders, leaders, and decision-makers at all stages of the development and implementation of AI applications.
5. Facilitate multistakeholder dialogue on how to advance AI innovation to increase trust and adoption and to inform future policy discussions.
6. Support efforts to promote trust in the development and adoption of AI systems with particular attention to countering harmful stereotypes and fostering gender equality. Foster initiatives that promote safety and transparency and provide guidance on human intervention in AI decision-making processes.
7. Promote the use of AI applications by companies, in particular small and medium-sized enterprises and companies from non-tech sectors.
8. Promote active labour market policies, workforce development and reskilling programs to develop the skills needed for new jobs and for those at risk of being left out, including policies specifically targeting the needs of women and underrepresented populations in order to increase labour participation rates for those groups.
9. Encourage investment in AI technology and innovation to create new opportunities for all people, especially to give greater support and options for unpaid caregivers, the majority of whom today are women.
10. Encourage initiatives, including those led by industry, to improve digital security in AI and developing technologies, such as the Internet of Things and cloud services, as well as through the development of voluntary codes of conduct, standards or guidelines and the sharing of best practices.
11. Ensure AI design and implementation respect and promote applicable frameworks for privacy and personal data protection.
12. Support an open and fair market environment including the free flow of information, while respecting applicable frameworks for privacy and data protection for AI innovation by addressing discriminatory trade practices, such as forced technology transfer, unjustified data localization requirements and source code disclosure, and recognizing the need for effective protection and enforcement of intellectual property rights.

## Annex 2 – AI Policy – Vodafone

Vodafone Group’s Artificial Intelligence (AI) Framework[[19]](#footnote-19) sets out our approach to working with AI technologies and

outlines how we intend to develop and employ it in a responsible manner across our international business; this also applies to the standards we expect from third parties developing AI systems in collaboration with and on behalf of Vodafone. We define AI as the application of advanced analytical techniques (such as Machine Learning, and Natural Language Processing) combined with automation to solve problems, develop personalised products and services, and seize opportunities in new ways.

At Vodafone, we are using AI to help to improve our products and services and to run our business as effectively as possible. For example:

* AI-powered chat bots increase the speed with which customer enquiries can be resolved;
* AI techniques in our networks are used to identify where capacity is needed so that our customers can enjoy optimised data services, such as high-quality video streaming;
* Vodafone employees use AI tools and software to help them work more efficiently; and
* We increasingly use AI to help support good decision-making, utilising ‘big data’ analysis based on large, anonymised data sets.

As AI grows in usage and impact across geographies and industries, Vodafone has a responsibility to consider how our use of this technology affects our customers, our employees, and wider society. We believe it is critical to ensure that the AI technologies we create, and employ are designed to respect the privacy and security of the end user’s data and their associated fundamental rights. The customer data we use is pseudonymised and permissioned. We will also seek to make AI-driven decisions that are fair and free of any harmful bias.

### Transparency and Accountability

We endeavour to clearly inform our customers and employees when they communicate directly with AI-powered systems.

AI-based systems, like intelligent chat bots, have the ability to seem increasingly human in their responses. We believe that people should be informed about when they interact with an algorithm or some form of AI/non-human system.

### Ethics and Fairness

We endeavour to develop AI in an ethical way so that it can be trusted.

State-of-the-art AI-based systems leverage large anonymised or pseudonymised data sets. It is critical that the outputs from these data-driven systems do not inadvertently guide us to make decisions that may affect any group or individual in an unfair way. Vodafone will strive to ensure that there is effective oversight and a ‘human-in-control’ approach to the use of AI. Beyond that, it is also important that Vodafone contributes to the debate about how this technology affects the societies we live in and is made as widely available as possible.

### Preservation of Privacy and Security

We endeavour to respect the privacy and protect the security of all individuals served by the AI we develop.

Customer trust is our number one priority. Respecting the privacy of our customers is essential to maintaining their trust in our business. Managing privacy risks effectively, including securing our network and putting customers in control of their data, is core to our approach.

### Human Rights, Diversity, and Inclusivity

We will ensure that we respect international human rights standards and best practice around ensuring AI systems foster diversity, accessibility, and inclusivity.

Vodafone upholds international human rights across its business footprint. Many of the issues that arise in the context of AI and human rights are not novel, but are exacerbated by the scale, proliferation, and real-life impact that AI facilitates. Because of this, the potential of AI to both strengthen and diminish human rights is much greater than in previous waves of technological development.

### Maximising the Benefits of AI While Managing the Disruption of its Implementation

Vodafone is a responsible employer and is determined to become a leading, human-centric, digital business.

Over time, we expect AI to automate an increasing number of routine tasks, enabling our employees to spend more time on higher value and rewarding activities, including the innovation needed to underpin sustainable business growth into the future.

## Annex 3 – The AI Audit framework

Table

Description automatically generatedAn 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[[20]](#footnote-20).

### 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.

# References

AI is a vast subject with impact on many areas and the same could be expected of AI within the IALA AtoN domain. Below are a few interesting links that could be explored.

## AI Auditing

[https://hbr.org/2018/11/why-we-need-to-audit-algorithms#](https://hbr.org/2018/11/why-we-need-to-audit-algorithms)  
<https://hellofuture.orange.com/en/auditing-ai-when-algorithms-come-under-scrutiny/>  
<https://www.informationweek.com/big-data/ai-machine-learning/how-well-conduct-algorithmic-audits-in-the-new-economy/a/d-id/1340299>

<https://venturebeat.com/2021/01/30/what-algorithm-auditing-startups-need-to-succeed/>  
<https://themarkup.org/ask-the-markup/2021/02/23/can-auditing-eliminate-bias-from-algorithms>  
<https://www.fastcompany.com/90597594/ai-algorithm-auditing-hirevue>

## AI Accuracy

<https://www.forbes.com/sites/ganeskesari/2021/01/21/accuracy-isnt-everything-how-even-a-wrong-ai-model-can-transform-your-business/?sh=57ce15ca7083>  
<https://wiki.pathmind.com/accuracy-precision-recall-f1>  
<https://developers.google.com/machine-learning/crash-course/classification/accuracy>  
<https://www.kdnuggets.com/2020/05/guide-choose-right-machine-learning-algorithm.html>  
<https://medium.datadriveninvestor.com/the-99-accurate-machine-learning-algorithms-you-shouldnt-buy-77fb6a86b436>

## The AI Black Box

<https://whatis.techtarget.com/definition/black-box-AI>  
<https://www.kdnuggets.com/2019/03/ai-black-box-explanation-problem.html>  
<https://www.thinkautomation.com/bots-and-ai/the-ai-black-box-problem/>  
<https://www.thinkautomation.com/eli5/eli5-explainable-ai/>

## AI Technology

<https://www.thinkautomation.com/eli5/deep-learning-vs-machine-learning-whats-the-difference/>

<https://longtermrisk.org/research-agenda#1_Introduction>

## ***AI Policies and strategies***

<https://futureoflife.org/national-international-ai-strategies/>  
<https://longtermrisk.org/coordination-challenges-for-preventing-ai-conflict/>

<https://www.ship-technology.com/features/ai-in-shipping/>

<https://www.oecd.ai/ai-principles>

<https://futureoflife.org/charlevoix-common-vision-future-artificial-intelligence/>

## AI and Intellectual Property

<https://link.springer.com/article/10.1007/s40319-020-00908-z>

# Action requested of the Committee

The Committee is requested to:

1. ........
2. Annex Heading 1
   1. Annex heading 2
      1. Annex heading 3
3. ........
4. Appendix heading 1
   1. Appendix heading 2
      1. Appendix heading 3

1. Input document number, to be assigned by the Committee Secretary [↑](#footnote-ref-1)
2. Leave open if uncertain [↑](#footnote-ref-2)
3. https://www.ibm.com/blogs/systems/ai-machine-learning-and-deep-learning-whats-the-difference/ [↑](#footnote-ref-3)
4. https://www.freecodecamp.org/news/want-to-know-how-deep-learning-works-heres-a-quick-guide-for-everyone-1aedeca88076/ [↑](#footnote-ref-4)
5. https://www.nature.com/articles/d41586-020-03186-4 [↑](#footnote-ref-5)
6. https://hdsr.mitpress.mit.edu/pub/f9kuryi8/release/6 [↑](#footnote-ref-6)
7. https://www.dataversity.net/thinking-inside-box-audit-ai/# [↑](#footnote-ref-7)
8. https://towardsdatascience.com/how-to-apply-continual-learning-to-your-machine-learning-models-4754adcd7f7f [↑](#footnote-ref-8)
9. https://blog.tensorflow.org/2021/01/ml-metadata-version-control-for-ml.html [↑](#footnote-ref-9)
10. https://thenewstack.io/add-it-up-how-long-does-a-machine-learning-deployment-take/ [↑](#footnote-ref-10)
11. https://www.technologyreview.com/2021/04/01/1021619/ai-data-errors-warp-machine-learning-progress/ [↑](#footnote-ref-11)
12. https://www.darkreading.com/vulnerabilities---threats/advanced-threats/expect-an-increase-in-attacks-on-ai-systems/d/d-id/1340833 [↑](#footnote-ref-12)
13. https://www.healthcareitnews.com/news/apac/ai-helps-radiologists-improve-accuracy-breast-cancer-detection-lesser-recalls [↑](#footnote-ref-13)
14. https://www.cigionline.org/articles/artificial-intelligence-and-keeping-humans-loop/ [↑](#footnote-ref-14)
15. https://www.genpact.com/insight/article/why-ai-still-needs-humans-in-the-loop [↑](#footnote-ref-15)
16. https://www.forbes.com/sites/forbestechcouncil/2019/10/30/why-ai-needs-human-input-and-always-will/?sh=352ae0135ff7 [↑](#footnote-ref-16)
17. https://www.chemistryworld.com/news/patent-applications-listing-ai-as-an-inventor-run-into-legal-problems/4013138.article [↑](#footnote-ref-17)
18. http://www.g7.utoronto.ca/summit/2018charlevoix/ai-commitment.html [↑](#footnote-ref-18)
19. https://www.vodafone.com/about-vodafone/how-we-operate/public-policy/policy-positions/artificial-intelligence-framework [↑](#footnote-ref-19)
20. https://doi.org/10.1145/3351095.3372873 [↑](#footnote-ref-20)