Quality Control in the Development Process of AI System on Ships

Hideki MIYAJIMA*

1. INTRODUCTION

Recently, the development of AI systems has advanced in the maritime industry. For example, a situation awareness support system [1) 2) 3)] using image recognition technology has been developed, and research and demonstration experiments [4) 5) 6)] on autonomous navigation AI that performs collision avoidance and automated berthing and unberthing have been carried out. At present, AI is mainly used to support seafarers or is in the research stage, but in the future, it is considered that AI systems installed on ships will be utilized in a way that is directly connected to safety.

This type of AI uses machine learning technology that learns patterns and rules from data. In addition, deep learning, which has attracted much attention recently, has shown high performance by learning complex models with large amounts of data. However, such AI has black box problem, and it is difficult for humans to understand how an AI system arrives at its conclusions. Also, AI system may output results that were not intended by the developer.

When considering AI systems from a quality assurance perspective, it is not easy to evaluate whether AI can perform as expected because AI with a mechanism in which its behavior is determined inductively by learning based on data. Furthermore, continuous learning is carried out in many cases so that the AI can follow changes in the surrounding situation even after development. Therefore, development considering the life cycle from design to operation is required. In the development of AI systems, it is important that activities to ensure quality, which varies depending on the purpose and use, are carried out throughout the development process. Considering such appropriate activities will lead to consideration of specific evaluation methods for AI systems.

Based on the development process and the importance of ensuring quality of AI systems, in order to ensure that AI systems developed under appropriate quality control are used in the maritime industry, Nippon Kaiji Kyokai (ClassNK, hereinafter, the Society) plan to develop “Technical Guide for Utilizing AI System on Ships -Quality Control in the Development Process of AI System-.” The items that AI system developers should consider during development will be described in this Guide. Therefore, as an introduction to this Guide, this paper presents an overview of AI and describes an approach for integrated development and operation. In addition, as an information that will be helpful when organizing the development process of AI systems in the maritime industry in the future, this paper introduces the development process organized in “Machine Learning Quality Management Guideline.”

2. ARTIFICIAL INTELLIGENCE

2.1 What Is AI?

The term AI was proposed at the Dartmouth Conference in 1956, and various studies have been conducted since then. Although various discussions have considered the definition of AI, there is no established definition, and it is understood as a broad concept such as “Programs that works in a manner similar to human thought processes, or information processing and technology that humans perceive as intelligent.” [7)]

Although AI includes various technologies such as exploration and inference algorithms and expert systems, the center of the recent AI boom is machine learning. Machine learning is one of the methods for analyzing data, in which a mechanism corresponding to human learning is realized by computers. Based on the calculation method (algorithm), the computer can discover patterns and rules from input data (training data), and by applying those patterns and rules to new data, it is possible to identify and predict new data.

Deep learning is one of the methods of machine learning, and has attracted special attention as the method that triggered the

[*] Research Institute, Research and Development Division, ClassNK

recent AI boom. Deep learning uses a neural network that emulates the mechanism of nerve cells in the human brain, and a feature of deep learning is that it has a multi-layered structure of that neural network.

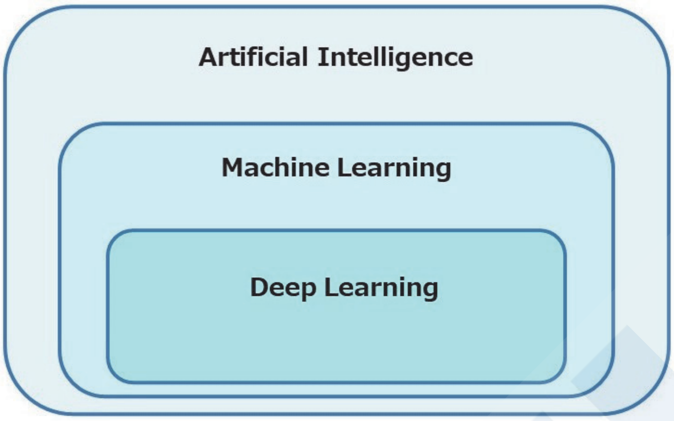

Fig. 1　Relationship between AI, machine learning and deep learning [7)]

2.2　Deep Learning

The neural network of deep learning consists of an input layer, an intermediate layer and an output layer. Features are extracted from the input data through calculations within the neural network, and the results of processing are output. The parameters for performing calculations of the neural network are optimized based on the data, and this process is called learning.

Deep learning achieves more accurate results than conventional methods in various fields by adopting a multi-layered network and learning with a large amount of data. There are various types of network structures, such as Convolutional Neural Networks (CNN) specialized in image recognition, Recurrent Neural Networks (RNN) used in processing time series data, text data, *etc*., generative AI which generates new data from a given data set, and so on, which are used depending on the purpose.

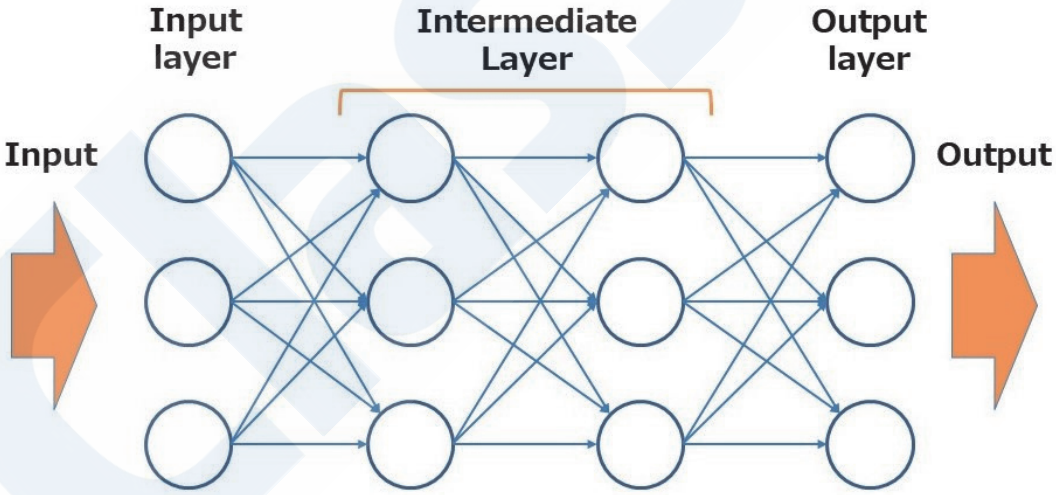

Fig. 2　Deep learning mechanism [8)]

2.3　Challenges of AI

AI based on machine learning can make predictions with high accuracy and is applied to a wide range of fields such as natural language processing, speech recognition, image recognition and content generation. On the other hand, since AI is learning based on a large amount of data and basically makes inferences probabilistically, it is necessary to deal with the following problems specific to AI.

2.3.1　Quality and Quantity of Data

Since AI learns based on data, the quality and quantity of the data are important for achieving the expected performance. For example, AI may not be expected to be accurate enough when past patterns do not apply to the data, the factors that serve as judgment criteria are not included in the data or the amount of data is small. Imbalanced data may also introduce errors in the inference results that the model outputs. Therefore, it is necessary to ensure that the data are adequate in terms of both quality and quantity. When developing an AI system, it is important to thoroughly analyze the problem to be solved, clarify the purpose, use environment and conditions, and then consider the quality and quantity of the data.

2.3.2 Computing Resources and Costs

Extensive data acquisition and annotation are extremely time-consuming. Large storage is also required to store and efficiently use big data. In addition, since a large amount of computational resources is required to train AI models, especially deep learning models, the time required for training must also be considered. Therefore, it is necessary to prepare a sufficient development system and environment.

2.3.3 Drift [9)]

In machine learning, a phenomenon called "drift" occurs in which the accuracy of predictions by the AI model decays over time. There are two main types of drift. With data drift, there is a discrepancy in the distribution of the data during training and the data during operation, such as seasonal changes and changes in trends. Conceptual drift refers to drift in which the relationship between the input data and the correct labels changes compared to when the model was trained, for example, when a new data feature appears. To deal with such drift, even after the development of AI systems, it is necessary to monitor changes in performance and the operating environment and to retrain the AI model periodically.

2.3.4 Probabilistic Behavior

AI basically makes probabilistic inferences but does not make decisions. It is also important to note that a phenomenon called "hallucination" [10)] occurs in generative AI, where AI generates information that is not based on reality. When developing an AI system, it is important to carry out activities to ensure that stakeholders fully understand these characteristics of AI, and to consider appropriate goals and operational procedures.

2.3.5 Black Box Problem

Machine learning models, especially deep learning, have a large number of parameters and a very complex structure. It is difficult for humans to understand how an AI system arrives at its conclusions, and the content of AI is essentially a black box. Although this kind of AI can perform prediction and recognition with high accuracy, the concern that "a system that cannot explain the basis of outputs cannot be used" may lead to distrust of the AI system, and may become a factor inhibiting its introduction. Therefore, it is important to carry out activities to gain the understanding from stakeholders, verify operational procedures considering the black box problem, and study the use of AI that can be explained rather than black box AI. XAI is attracting attention as a solution.

2.3.6 Difficulty of Evaluation

AI that learns based on data cannot explicitly relate its internal design and implementation to a defined specification, and it is difficult to review the internal design and implementation and evaluate quality. In addition, AI technology is advancing rapidly, and there are also various ways to utilize it. Therefore, it is important to clarify the objectives to be achieved in the whole system and the elements required in the AI system, and to consider the evaluation method according to the target AI system.

2.3.7 Differences in Risk by Utilization Method

The extent to which AI risks need to be considered varies greatly depending on how the AI system is used. For example, AI system is used in daily life, for example, in translation, chatbots and speech recognition, may have no risks which are paid attention. On the other hand, in the case of AI systems that are directly linked to safety, such as autonomous driving and autonomous ship navigation, safety and reliability should be carefully considered.

The AI Act, which is a European regulation on AI that was proposed by the European Commission on April 21, 2021, adopts a risk-based approach, establishing 4 risk levels and prescribing requirements and regulations according to each risk [11)]. This approach is useful for developing AI systems. It is important to grasp the risks of AI systems through risk assessment and to ensure safety by taking appropriate control measures.

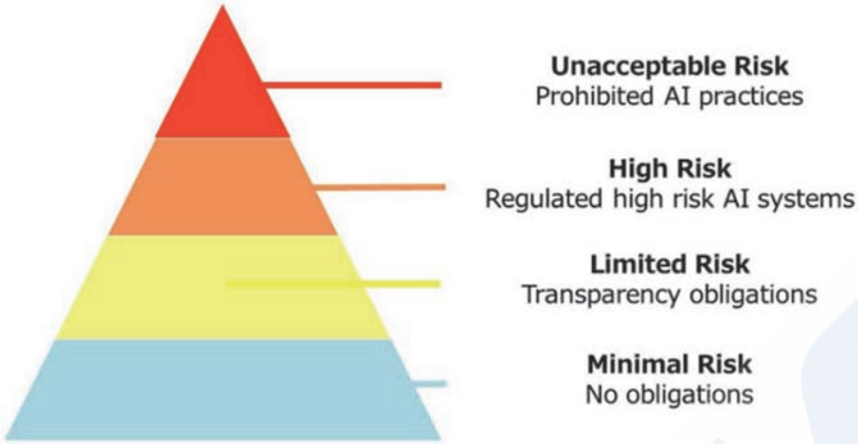

Fig. 3 Pyramid of risks in EU AI Act [12)]

2.3.8 Security [13)]

When developing AI systems, in addition to the conventional IT and system security viewpoints, it is also necessary to consider attacks targeting training data and AI models as security problems specific to AI. Attacks on AI models include the following. When developing AI systems, it is necessary to carry out a risk assessment and take countermeasures against attacks.

・Poisoning Attack

A type of attack in which the training data or AI model is manipulated in some way to cause the system to output results not intended by the AI model developer.

・Evasion Attack

Addition of noise to inference data can lead to incorrect inferences by AI models. When the added noise is small, a human may not be able to distinguish the difference from the original data. A kind of attack to discover such small noise is called an “evasion attack,” and data with small added noise is called an “adversarial example.”

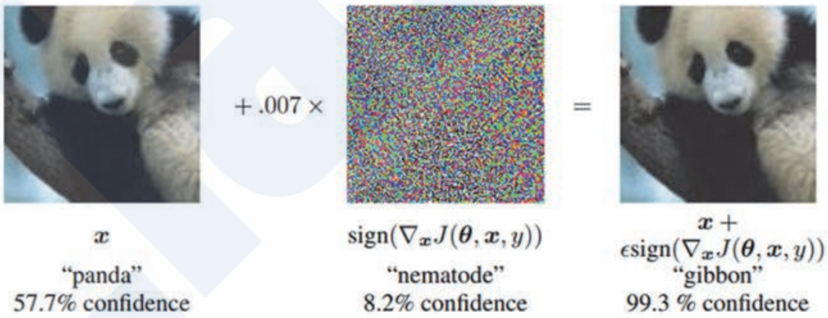
+ .007 ×
=
$\boldsymbol{x}$
“panda”
57.7% confidence
$\mathrm{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$
“nematode”
8.2% confidence
$\boldsymbol{x} + \epsilon\mathrm{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$
“gibbon”
99.3 % confidence

Fig. 4 Example of an adversarial example of a panda recognized as a gibbon [14)]

When developing AI systems, it is necessary to address these challenges through activities in the development process.

3. APPROACH FOR INTEGRATED DEVELOPMENT AND OPERATION

In general, AI systems are developed by repeatedly learning and evaluating models based on data to achieve the expected performance. The performance of the developed AI system is also monitored even after it is introduced into the real environment, and continuous evaluation and learning are performed to cope with changes in data trends caused by changes in the surrounding environment such as drift, and to improve accuracy using newly acquired data. Therefore, in the development of AI systems, an approach for integrated development and operation [15) 16)] is often adopted, in which quality is maintained and improved by repeating learning with newly acquired data so that performance approaches the expected result more closely. Considering these points, the Guide will organize the approach of the AI system from the beginning of development to the end of operation as shown in Fig. 5.

Since the actual development flow varies depending on the developer and the system to be developed, a general approach considering various AI systems is shown here. In the development of AI systems, through each phase of such an approach, it is required to carry out the activities necessary to ensure quality while addressing the aforementioned AI-specific challenges. Therefore, the items to be considered during development in each phase will be described in the Guide.

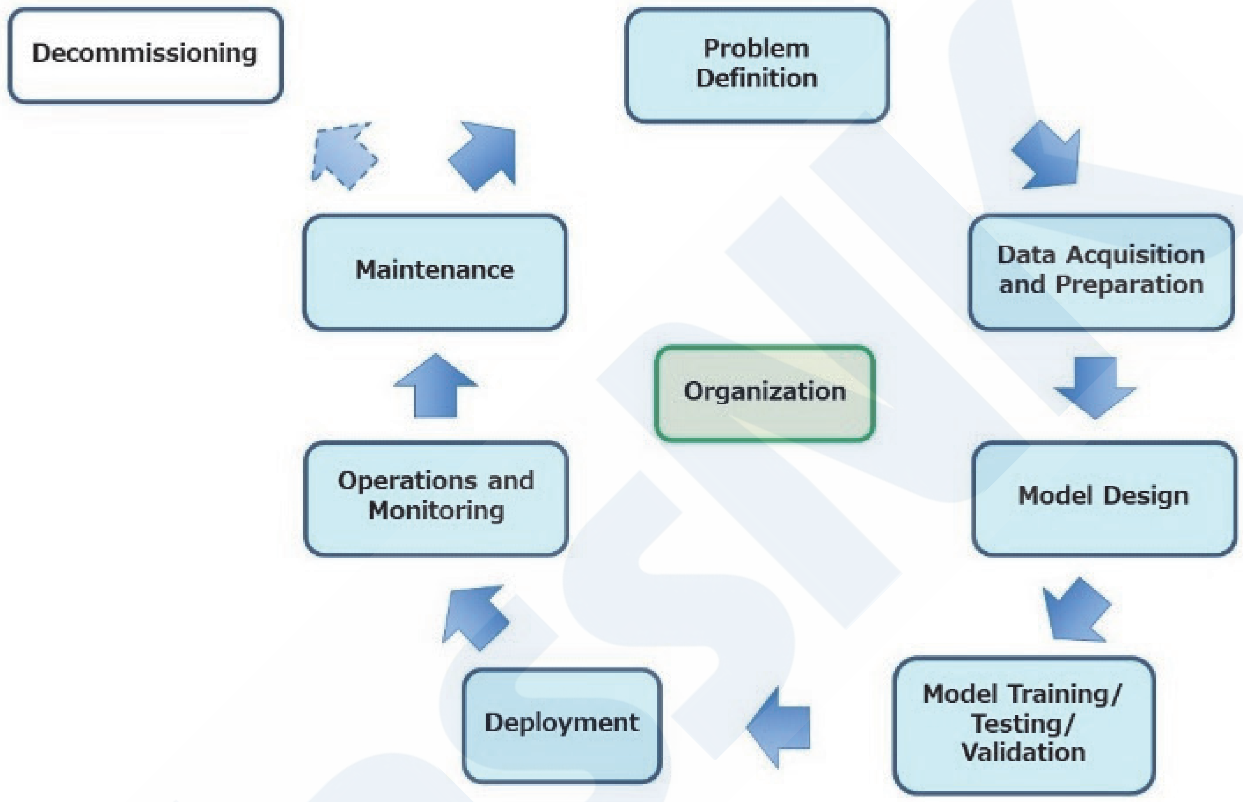
Decommissioning
Problem Definition
Data Acquisition and Preparation
Maintenance
Organization
Operations and Monitoring
Model Design
Model Training/ Testing/ Validation
Deployment

Fig. 5 Approach for integrated development and operation

In the development of an AI system, the development system and the development environment are also important, and since they are related to the whole approach, they are placed at the center of the organization.

0) Organization
Factors related to the ability of the organization involved in the development and operation necessary for the successful development of the AI system.

1) Problem Definition
The phase in which the problems to be solved by the AI system are identified and the necessary considerations are made to proceed with the development.

2) Data Acquisition and Preparation
The phase in which data are acquired and preprocessed for use in training the AI model.

3) Model Design
The phase in which the overall approach of development is defined and the AI model is designed based on the objectives, requirements and available data defined in the Problem Definition.

4) Model Training/Testing/Validation
The phase in which the AI model is trained and verified to function as expected.

5) Deployment
The phase in which the AI system is integrated with existing systems, processes, products and services so as to function

in a real-world environment.

6) Operations and Monitoring
The phase in which the AI system is operated in the real-world environment and is continuously monitored to ensure that the expected performance is being achieved.

7) Maintenance
The phase in which maintenance and system and model updates are performed as needed.

8) Decommissioning
The phase in which the operation results of the AI system are evaluated, and the operation of the AI system is terminated when the AI system is no longer needed or when another effective solution is developed.

4. ORGANIZING THE DEVELOPMENT PROCESS IN THE "MACHINE LEARNING QUALITY MANAGEMENT GUIDELINE"

As mentioned above, the approach for integrated development and operation can be used universally regardless of the target, and it is assumed that quality will be improved by continuously repeating learning. When this approach is applied to AI systems installed on ships, it must be reduced to a form suitable for the development process of the maritime industry. This paper introduces the development process modeled in "Machine Learning Quality Management Guidelines" as an idea which can be used in such cases.

In the development of an AI systems, the quality of the AI system is ensured through the approach which integrates development and operation described above. On the other hand, since AI learning depends on data, it is difficult to predict the deliverables in the early stage of development, and differences in understanding tend to occur between the developer and user. Therefore, PoC (Proof of Concept) is often performed when developing AI systems. PoC is a verification process that shows the feasibility of new concepts and ideas. In addition, even after an AI system is developed and introduced into a real environment, learning using newly acquired data is performed continuously to maintain and improve accuracy.

In the "Machine Learning Quality Management Guidelines" published by Japan's National Institute of Advanced Industrial Science and Technology, the development process of such general AI systems is divided into three stages, as shown in Fig. 6: "Proof of Concept stage," "Iterative development stage" and "DevOps and cont. training stage." it is important in the maritime industry to verify new technologies in advance of introduction and at the completion of development, and to confirm changes after the start of operation. Thus, it can be thought that the development stages described in the Guidelines are highly compatible with the maritime industry.

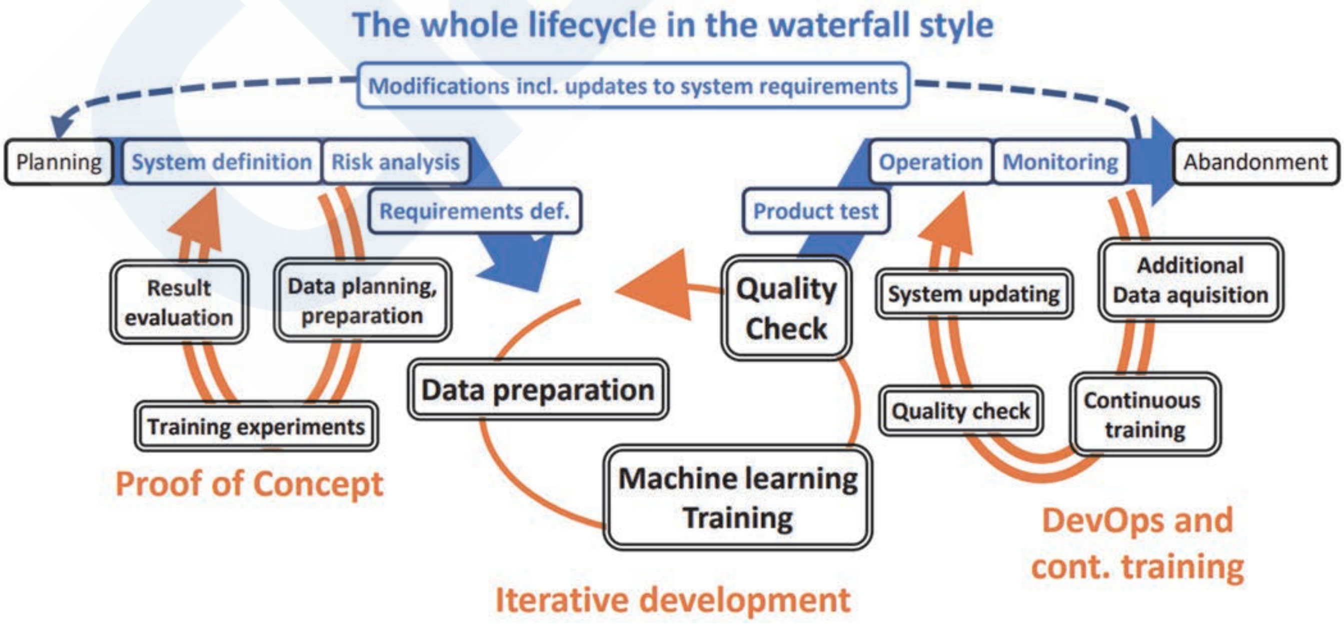

Fig. 6 Development process in "Machine Learning Quality Management Guideline" [15)]

5. CONCLUSION

Regarding "Technical Guide for Utilizing AI System on Ships -Quality Control in the Development Process of AI System-," which is under development, this paper has presented an overview of AI and described an approach for integrated development and operation. As a reference when applying this approach to the development process of the maritime industry, the development process in "Machine Learning Quality Management Guideline" was also introduced. Since the contents of the two guidelines can be used universally regardless of the application of the AI system, it will be necessary to arrange them in a form suitable for the development process in the maritime industry in future.

In the development of AI systems in the maritime industry, there will be cases where "people in the maritime industry develop AI systems" and cases where "companies responsible for AI system development develop AI system for the maritime industry." For the former, it is necessary to understand the AI system development process, and for the latter, it is necessary to understand the development process of the maritime industry. The Guide will strive to provide information that will provide insights in both cases. It is our hope that these contents will contribute to the development of AI systems in the maritime industry.

REFERENCES

1) JRCS: infoceanus, https://infoceanus.com/, 2024.3
2) GROKE: Groke Pro, https://www.groke-tech.com/en/products, 2024.3.
3) Orca AI: SeaPod, https://www.orca-ai.io/seapod/, 2024.3
4) The Nippon Foundation: The Nippon Foundation MEGURI2040 Fully Autonomous Ship Program Aiming to implement fully autonomous navigation to support the ocean of the future, https://www.nippon-foundation.or.jp/en/what/projects/meguri2040, 2024.3
5) The Nippon Foundation: The Nippon Foundation MEGURI2040 Fully Autonomous Ship Program Successful fully autonomous navigation of large, high-speed, coastal ferry in northern Kyushu seen leading to improved safety, https://www.nippon-foundation.or.jp/en/news/articles/2022/20220118-66716.html, 2024.3
6) Hashimoto et.al: Development of AI-based Automatic Collision Avoidance System and Evaluation by Actual Ship Experiment, ClassNK Technical Journal, 2021
7) Ministry of Internal Affairs & Communications: 2019 WHITE PAPER Information and Communications in Japan (in Japanese), https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r01/html/nd113210.html, 2024.3
8) Fumitaka Sato: Overview of Deep Learning ― The State-of-the-Art Machine Learning Technologies at the Core of AI ― (in Japanese), Fujitsu Journal, 2018
9) Masahiko Isshiki: What is Concept drift/Data drift? (in Japanese), https://atmarkit.itmedia.co.jp/ait/articles/2202/21/news033.html, 2024.03
10) AI White Paper Editorial Board: Artificial Intelligence White Paper 2023 Impact of Generative AI and AI Governance (in Japanese), Kadokawa Ascii Research Laboratories, 2023
11) PwC Japan: The forefront of regulatory trends in the US, Europe, and China regarding generative AI (in Japanese), https://www.pwc.com/jp/ja/knowledge/column/awareness-cyber-security/generative-ai-regulation03.html, 2024.03
12) Tambiama Madiega, et al.: Artificial intelligence act, European Parliament Research Service, 2023
13) Information-technology Promotion Agency: AI Handbook for Security Stakeholders (in Japanese), 2022
14) Ian J. Goodfellow et al.: Explaining and Harnessing Adversarial Examples, International Conference on Learning Representations ICLR, 2015
15) National Institute of Advanced Industrial Science and Technology: Machine Learning Quality Management Guideline, 1st edition, 2021
16) ISO/IEC 22989: 2022 Information technology — Artificial intelligence — Artificial intelligence concepts and terminology