Guidelines on Assessment of AI Reliability in the Field of Plant Safety

Second Edition

March 2021

the Liaison Council of Three Ministries on Disaster Prevention of Petroleum Complexes

(Ministry of Economy, Trade and Industry; Fire and Disaster Management Agency; and Ministry of Health, Labour and Welfare)

Executive Summary

Background

Petroleum and chemical plants are faced with aging facilities and a shortage of safety personnel, with a concern that the sustainability of their safety systems will decline.

On the other hand, new technologies such as the Internet of Things (IoT), drones, and artificial intelligence (AI) have become increasingly practical. Appropriate use of these new technologies to ensure stable plant operation will not only maintain and improve safety, but also stabilize product quality, improve cost efficiency, and meet construction and maintenance deadlines.

In particular, safety systems incorporating AI are being constructed in the field of plant safety in line with the recent technological development of machine learning. For example, there is AI that can detect signs of a slight abnormality based on the relationship among a large amount of sensor data and AI that can optimize the operation of distillation equipment, etc. to increase productivity. They are being demonstrated to dramatically enhance safety and productivity.

Challenges

In order to proceed from demonstration to implementation, it is necessary to appropriately verify that the AI performs as expected in quality (i.e. reliability).

However, a methodology for AI reliability assessment has not yet been established, and this is one of the main reasonsfor limited introduction of AI in the field of plant safety, where safety is crucial.

Reference: Comments from business operators on the necessity of AI reliability assessment in the field of plant safety

- Huge loss (human injury and economic loss) will result if the operation of the plant is interrupted because of AI defects. Therefore, in order to gain understanding from relevant departments in the company, a high level of reliability assessment is necessary, and it is currently difficult to achieve. (Plant owner)
- Since reliability has not been sufficiently assessed, we cannot leave the management of important facilities to AI, and we will leave only non-critical facilities to AI. On the other hand, unimportant facilities may not need to be maintained in the first place, and inspections themselves may be unnecessary. (Plant owner)
- As an AI vendor, we have a hard time getting the customers (plant owners) to understand the reliability of our AI. (AI vendor)

In recent years, advanced studies have been conducted in Japan on quality assurance and assessment, including AI reliability. The National Institute of Advanced Industrial Science and Technology (AIST) (2020) ”Machine Learning Quality Management Guideline 1st edition” and Consortium of Quality Assurance for Artificial-Intelligence-based products and services, (QA4AI consortium) (2020) " Guideline for quality assurance of AI-based products 2020.08" formulate the methodlogy of reliability assessment and points to note. However, these guidelines are cross-sectoral in nature, including e-commerce and automated driving, and do not specifically consider reliability assessment in the field of plant safety.

Therefore, in order to promote implementation of AI with high reliability in the field of plant safety, it is necessary to organize how to interpret and apply the cross-sectoral methodology of reliability assessment of AI in the field of plant safety.

Purpose and application of the Guidelines

Recognizing the above issues, the Guidelines show how to properly manage AI reliability (fulfillment of expected quality for plant safety and productivity improvement) dedicated to the field of plant safety.

By using the Guidelines, plant owner companies can achieve highly reliable AI and improve safety and productivity. It will also make it easier to be accountable internally and externally for the reliability of AI. Furthermore, when developing the system with vendor companies, the requirements can be communicated appropriately, and the achievement status can be confirmed smoothly.

A vendor company that develops and delivers AI can more easily explain the reliability of AI to the plant owner company because the Guidelines clarify how to build AI with sufficient reliability. In addition, it is expected that the Guidelines will allow for appropriate setting of requirements in the AI development process with the owner company.

Structure and flow of reliability assessment

Basic Concept

- In the Guidelines, the reliability assessment of AI is conducted in the same way as the National Institute of Advanced Industrial Science and Technology (AIST) (2020) "Machine Learning Quality Management Guideline 1st edition" and the measures to ensure the necessary reliability are presented. Then, based on application examples of AI in the field of plant safety, specific applications to the field of plant safety are shown.
- The Guidelines focus on reliability assessment of machine learning, which has been put to practical use in recent years.

The quality of machine learning is divided into three levels ("quality in use," "external quality," and "internal quality," in descending order of hierarchy), and the reliability of the machine learning (ML) user system is managed by achieving these levels (See ch. 2).

“Quality in use": The quality required to be achieved by the whole system including ML components (See ch. 2.1.1).

“External quality": The quality that must be met in the ML components in order to satisfy "quality in use" (See ch. 2.1.2). It is classified into two types: "risk avoidance" and "performance."

- Risk avoidance: A type of quality that pursues safety. The objective is to avoid or reduce the risk of adverse effects (human suffering and economic damage) due to misjudgment of ML components (See ch. 2.1.3 (1)).
- Performance: A type of quality that pursues productivity. The objective is to perform plant operations and inspections efficiently (See ch. 2.1.3 (2)).

“Internal quality”: The quality that must be met in the design, development, operation, etc. of ML components in order to satisfy external quality. Eight internal qualities are defined for the following three types (See ch. 2.1.4).

- Appropriateness of data used in development (e.g. amount and type of data)
- Appropriateness of the model developed (e.g. accuracy of machine learning during testing)

- ➢ Appropriateness of implementation and operation methods (e.g. how to maintain the accuracy of machine learning)

The basic procedure for reliability assessment is as follows (Figure 1).

(1) Set the "quality in use" (See ch. 2.2.1).

(2) Set the "external quality" required to achieve "quality in use" and determine the level of achievement (level of demand) of the "external quality" for each classification of "risk avoidance" and "performance.” There are four levels for risk avoidance (Levels 0, 0.1, 0.2, and 1) and three levels for performance (Levels 0, 1, and 2). Levels for the external quality are to be set, in accordance with the procedure described in the Guidelines(See ch. 2.2.2, 2.2.3).

(3) The required level of "internal quality" (Level 1, 2, or 3) is determined by the set level of "external quality." (For example, the higher the level of the internal quality regarding the type and amount of training data, the more diverse and larger amount of data must be collected.) Machine learning components are developed accordingly (See ch. 2.2.4, 2.2.5).

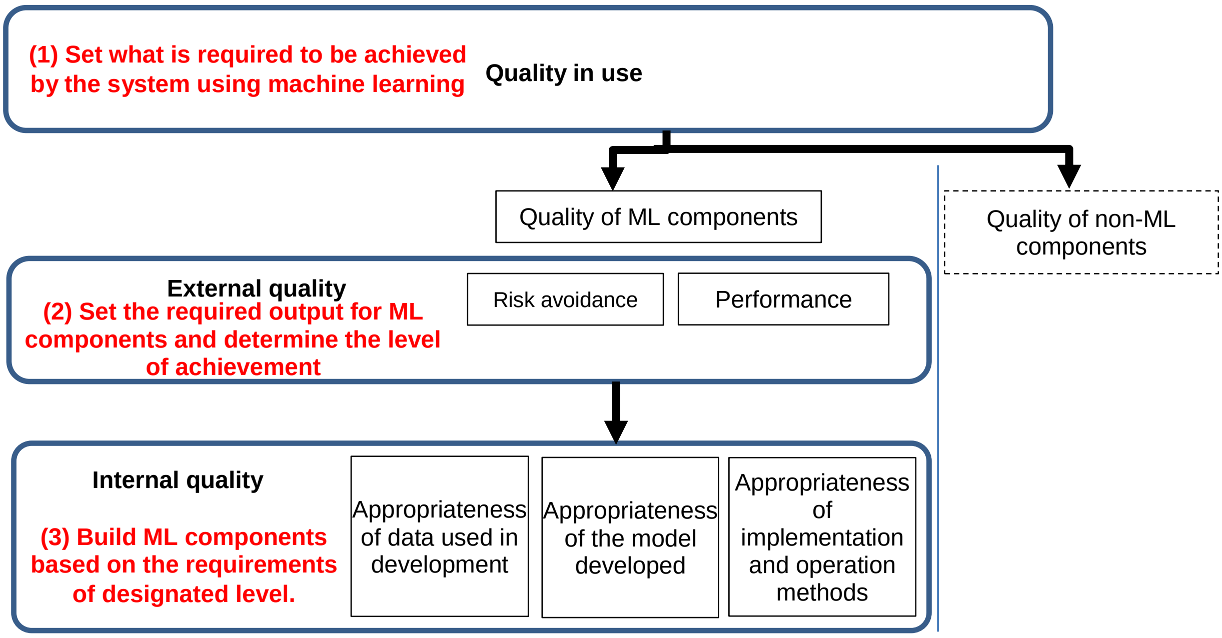


Figure 1 Procedure for reliability assessment

Taking the example of "detection and diagnosis of early signs of abnormality" in the Guidelines (machine learning in which ML components analyze a large number of sensor data to detect signs of abnormalities in operation that will become apparent in 20–30 minutes to a few days), which is discussed in the Guidelines as an example of a use case in the field of plant safety (See ch. 3.3.4), reliability is ensured through the following flow (Figure 2).

(1) Verbalize what needs to be achieved in this case and specify the "quality in use."
(2) Specify the "external quality" of the ML components corresponding to the "quality in use" and determine the level of "risk avoidance" and "performance" according to the procedures described in the Guidelines. In the example in Figure 2, risk avoidance is tentatively set to level 0.1 and performance to level 1.
(3) Level 0.1 for risk avoidance and level 1 for performance correspond to level 1 of "internal quality." Level 1 requirements and the points to be considered in executing them are confirmed and ML components are developed.

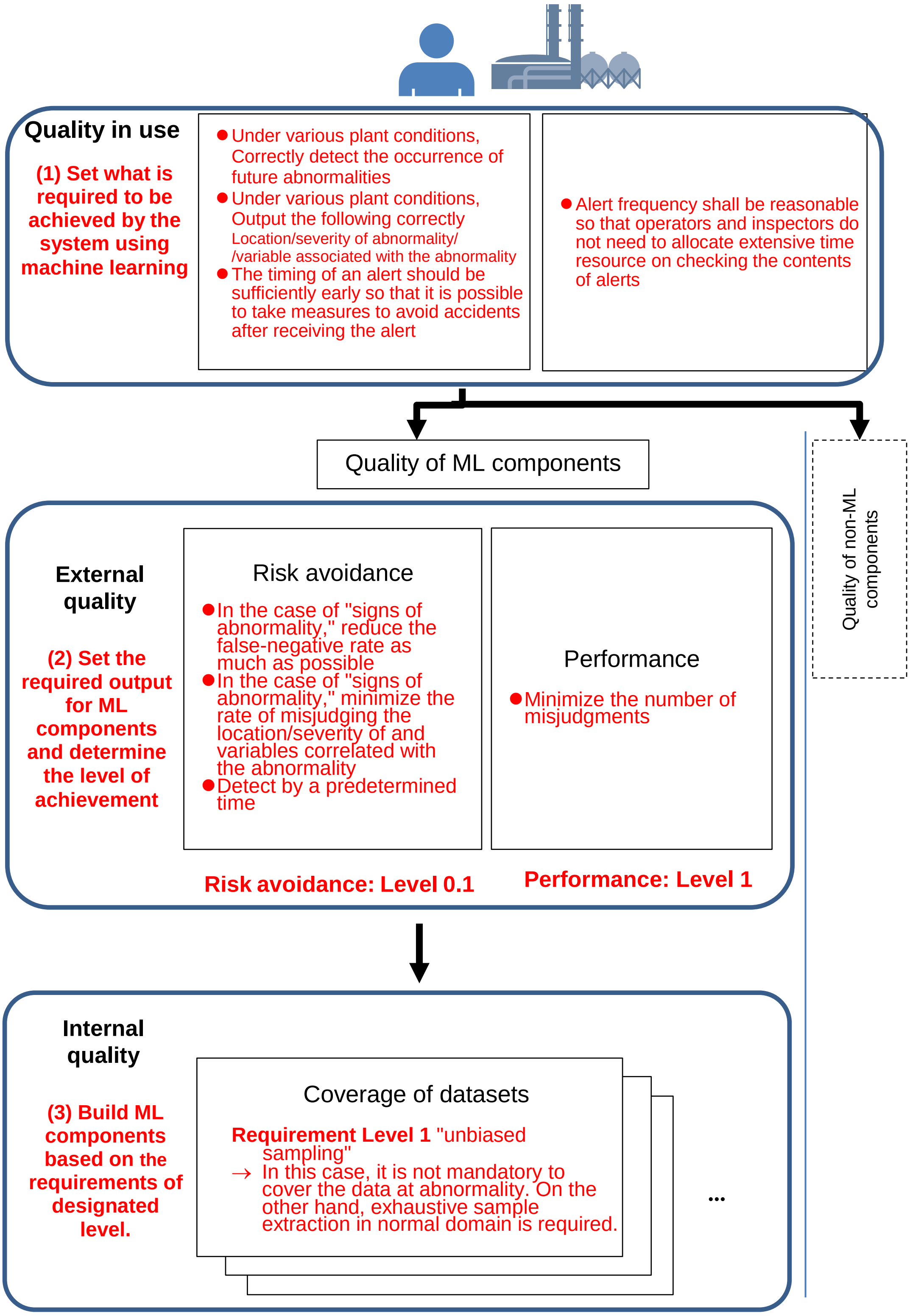


Figure 2 Flow of reliability assessment (using the "detection and diagnosis of early signs of abnormality" as an example)

Contents specific to the field of plant safety

In addition to the above-mentioned "detection and diagnosis of early signs of abnormality," the Guidelines also cover "optimization of operation," "prediction of pipe wall thickness," "pipeline image diagnosis" and "equipment deterioration diagnosis" as use cases. Specific examples of quality in use and external quality items, and points to keep in mind when executing internal quality requirements are provided (See ch. 3).

If there is a use case similar to the ML components being developed by one’s own company, it is assumed that the examples in the Guideline will be used as a reference. Even if there is no similar use case, the structure and flow of reliability assessment can be applied as is. It is expected that the items and levels of quality in use and external quality will be considered according to the case of one’s own company.

In addition, how the Guidelines should be utilized by each personnel related, e.g. plant system staff, quality assurance staff, maintenance engineers etc., is shown in accordance with the specific flow of the Guidelines (See ch. 4).

Furthermore, in order to ensure that the implementations described in the Guidelines are achieved without fail, a template recording the results of reliability assessment process is also available[1]. By recording the results of assessment according to the template, evidence to which the Guidelines are applied can be kept and used for explanations within and without the company. In addition, records of actual reliability assessments using the template are available as “Practical Examples” for all use cases. Consulting Practical Examples similar to one’s own situation would yield appropriate level of detail to be written in the template.

[1]https://www.fdma.go.jp/relocation/neuter/topics/fieldList4_16/jisyuhoan_shiryo.html

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1 Purpose and Application of the Guidelines

1.1.1 Purpose of the Guidelines

The National Institute of Advanced Industrial Science and Technology (AIST) ”Machine Learning Quality Management Guideline 1st edition” describes the potential and obstacles to industrial applications of AI as follows:

" The effectiveness of artificial Intelligence (AI), especially machine learning technology, has been accepted in broad fields of applications such as manufacturing, automated driving, robots, health care, finance and retail business, and its social implementation seems to start to blossom. ***On the other hand, <u>it is difficult to identify the cause when any accident occurs or to explain advantages of AI-based products to the amount of investments due to the lack of technologies to measure and demonstrate the quality of AI-based products or services. Consequently, wider acceptance of AI in the society is lagging, causing a big obstacle to the expansion of AI development business</u>.*** *”*[2]

In particular, the field of plant safety is characterized by the facts that “(1) the function that AI supports and replaces is safety (i.e. protective functions),” and that “(2) it requires accountability to a diverse set of stakeholders.” For these reasons, the issue of the lack of technology to account for the reliability of AI is prominent.

As for (1), the safety function has been performed by humans and existing systems and software, and a structure[3] has been developed to secure and assess the reliability of such systems to operate properly and perform necessary functions. In this way, the safety function has ensured the public safety and the safety of workers. In order to incorporate AI into the safety function as a support or substitute for the function, an established methodology for ensuring and assessing the reliability of AI behavior is required. Without such methodology, AI will be incorporated into safety functions without a clear rationale for ensuring safety, and safety will be compromised.

Regarding (2), plant owner companies need to first have appropriate awareness of the probable effects and safety risks of AI within the company, including the environment and safety departments and the management team, and only then agree on the introduction of AI. In addition, it is necessary to explain the safety activities to the local communities and regulatory authorities outside the company. When doing so, it is necessary to show evidence that AI is being developed, implemented, and operated in an appropriate process, and to build consensus.

Table 1-1 shows the perception of business operators regarding the reliability of AI in plant safety.

[2] National Institute of Advanced Industrial Science and Technology (AIST) (2020), “1st edition of Guidelines on Quality Control for Machine Learning”

[3] Security assurance measures specified by laws and regulations, international standards related to functional safety and software quality, etc.

Table 1-1 Business operators' view of AI reliability issue in plant safety[4]

➢ [Related to (1)] Since reliability has not been sufficiently assessed, we cannot leave the management of important facilities to AI, and we will leave only non-critical facilities to AI. On the other hand, unimportant facilities may not need to be maintained in the first place, and inspections themselves may be unnecessary. (Plant owner) ➢ [Related to (1)/(2)] Huge loss (human injury and economic loss) will result if the operation of the plant is interrupted because of AI defects. Therefore, in order to gain understanding from relevant departments in the company, a high level of reliability assessment is necessary, and it is currently difficult to achieve. (Plant owner) ➢ [Related to (2)] For the Super Certification, it is necessary to explain the appropriateness of safety measures to the authorities, etc. For this reason, it takes a very long time before a decision is made within the company that it is acceptable to use AI when conducting a safety inspection that deviates from KHKS[5]. (Plant owner) ➢ [Related to (2)] As an AI vendor, we have a hard time getting the customers (plant owners) to understand the reliability of our AI. If guidelines for reliability assessment are established, customers will be satisfied by referring to them. (AI vendor)

Thus, in the field of plant safety, there is a particular need for a systematic methodology to assess the reliability of AI (i.e. the challenge of the Guidelines). In order to solve this problem, the Guidelines aim to present the concept of assessing the reliability of machine learning, which has become increasingly practical in AI, dedicated to the field of plant safety.

1.1.2 Benefits of using the Guidelines

The main readers of the Guidelines are assumed to be plant owner companies that install machine learning systems in plants (hereinafter referred to as "ML based system"), and vendor companies that develop and deliver ML based system. The following describes the expected benefits from each position assumed.
Chapter 4 of the Guidelines describes in detail the situations in which the Guidelines are used and the reference points for each stakeholder involved in the process of developing and operating the ML based system.

(1) Benefits for plant owner companies

By following the reliability assessment method presented in the Guidelines, appropriate safety and productivity requirements can be set for software components (hereinafter referred to as "ML components") built using machine learning technology in the safety system, and specific requirements can be met to achieve them. Through this process, high reliability of ML components can be achieved with improved safety and productivity.
In addition, by providing evidence of the development and operation of ML components in accordance with this document, it becomes easier to have accountability for the safety, etc. of ML components

[4] Based on the interview survey conducted in the development of the Guidelines.
[5] Safety inspection standards established by the High Pressure Gas Safety Institute of Japan. It is specified as a method of safety inspection in the "Notification stipulating the Method of Safety Inspection" (Ministry of Economy, Trade and Industry Notification No. 84, March 30, 2005). However, in the case of "Specified Authorized Business Operators (Super Certified Business Site)," if a business operator judges that the method is sufficient to check the status of damage, deformation, and abnormality, etc., the business operator may use a method freely set by themselves instead of these safety inspection standards.

both within and without the company. Furthermore, when developing the safety system with vendor companies, the requirements can be communicated appropriately, and the achievement status can be confirmed smoothly.

(2) Benefits for vendor companies

The vendor companies can use the Guidelines as a common language for the development and operation of machine learning with the plant owner companies. Specifically, the vendor companies can receive various requirements from the plant owner companies for ML components by converting them into the levels set in the Guidelines. In addition, when developing ML components, simply implementing the requirements set forth in this document corresponding to the levels would be sufficient, absolving the vendor companies of the complexness in explaining the validity of their actions and methodology to plant owner companies.
This aids the vendor companies to gain understanding from the plant owner companies about the reliability of ML components. Furthermore, it allows them to differentiate their service from those that do not perform appropriate reliability assessment.

1.1.3 Background to the formulation of Guidelines

From April 2020 to March 2021, the "Working Group on AI Reliability Assessment in Plants[6]" was held by operators in the field of plant safety and safety and AI experts. The working group discussed the status of machine learning in the field of plant safety and the methods of reliability assessment based on the study results. The Guidelines have been formulated based on the results of this study.
In the future, the Guidelines will be continuously reviewed based on the status of their use, technological development in the field of plant safety, and progress in AI and its quality assurance and assessment technologies.

[6] See Annex of the Guidelines for the composition of study group members, etc.

1.2 Relationship with other Guidelines

1.2.1 Existing guidelines related to AI quality

No methodology has been established in the world for quality assurance and assessment including the reliability of machine learning, but advanced studies are being conducted in Japan. In developing the Guidelines, the following guidelines have mainly been used as reference.

(1) National Institute of Advanced Industrial Science and Technology (AIST), “Machine Learning Quality Management Guideline 1st edition”

The National Institute of Advanced Industrial Science and Technology (AIST) ”Machine Learning Quality Management Guideline 1st edition” is "a systematic analysis of the efforts and inspection items needed to satisfy the quality requirements of service provision"[7] for ML based system. The abovementioned guidelines categorize the quality of ML based system into three levels, "quality in use," "external quality" and "internal quality"[8]. They organize them as "achieving the necessary level of 'external quality' by improving the 'internal quality' of ML components and realizing the 'quality in use' of the final product"[7]. The quality demands for ML components (external quality) are divided into three axes: risk avoidance, AI performance, and fairness. Then the level of each axis is set according to the level of demand. In addition, the requirements related to the development process and data that should be implemented when developing ML components according to the quality level are organized. The ”Machine Learning Quality Management Guideline 1st edition” is universal in nature, not intended for any particular industry, and will require the development of industry-specific references for application to specific fields.

(2) Consortium of Quality Assurance for Artificial-Intelligence-based products and services"Guideline for quality assurance of AI-based products 2020.08”

The "Guideline for quality assurance of AI-based products 2020.08" is a “common guideline for quality assurance of AI products”[9] developed by the " Consortium of Quality Assurance for Artificial-Intelligence-based products and services," which consists of individuals and organizations from private companies, universities, and research institutes. As a framework for quality assurance of AI products, it presents the five axes of Data Integrity, Model Robustness, System Quality, Process Agility, and Customer Expectation, and organizes the items to be considered in each axis in the form of a checklist. In addition, it organizes a catalog of technologies that are useful for quality assurance and provides examples of domain-specific guidelines. The five domains illustrated are: content generation systems, smart speakers, industrial processes, automated driving, and AI-OCR.

1.2.2 Positioning of the Guidelines

The Guidelines adopt the system of the ”Machine Learning Quality Management Guideline 1st edition” and are positioned as a reference that shows how to apply them to the field of plant safety. Figure 1-1

---

7 “AIST Announces Guidelines on Quality Control for Machine Learning," https://www.aist.go.jp/aist_j/press_release/pr2020/pr20200630_2/pr20200630_2.html, accessed on September 18, 2020

8 See Chapter 2 for details.

9 Consortium of Quality Assurance for Artificial-Intelligence-based products and services "Guideline for quality assurance of AI-based products 2020.08”

shows the relationship with the ”Machine Learning Quality Management Guideline 1st edition”.
The Guidelines use the reliability assessment structure of the ”Machine Learning Quality Management Guideline 1st edition”. Specifically, the two have identical structure in terms of the three-level hierarchy of "quality in use," "external quality," and "internal quality," two axes and levels for external quality, and eight axes and requirements for internal quality (left half of Figure 1-1).
Furthermore, the structure of reliability assessment is fleshed out based on actual case studies in the field of plant safety. Specifically, it shows concrete examples for quality in use and external quality, procedures for setting the level based on the safety function of the entire system, and specific perspectives for realizing requirements (right half of Figure 1-1).

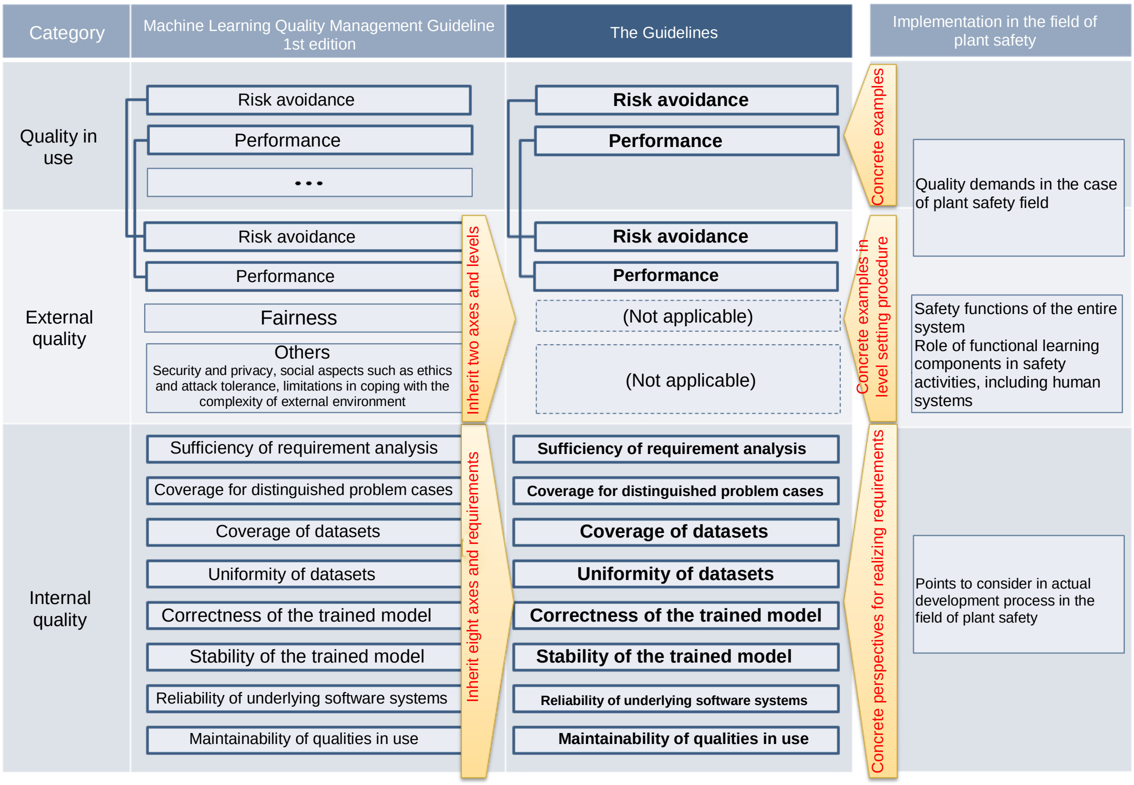


Figure 1-1 Relationship between the Guidelines and the AIST Guidelines

In addition, Chapter 7 of the “Guideline for quality assurance of AI-based products 2020.08” describes quality assurance in the "industrial process" domain, which is about applying ML components to plant controls. The points listed here have been added as "Concrete perspectives for realizing requirements" (detailed in 2.1.4).

1.3 Scope of Application

The Guidelines thoroughly analyses the reliability of ML components [10]and lists the issues that should be considered from the perspective of reliability when applying machine learning to the field of plant safety, in addition to the conventional actions to ensure reliability of plant safety such as functional safety[11].

For example, the <u>risk</u> addressed in the Guidelines is the danger and disaster caused by misjudgment of <u>ML components</u>. The Guidelines do not include hardware reliability and dangerous failures, and do not cover danger and disaster caused by failures of hardware that executes processing. In addition, the Guidelines do not address the reliability of software other than ML components, nor do they address the operations (procedures, manuals, etc.) of personnel responsible for plant safety.

In other words, the safety and performance of the entire system including ML components and the entire plant cannot be achieved solely by consulting these Guidelines.

In addition, legal and ethical issues, privacy of third parties, social acceptance, and cyber security, which are generally considered to be issues in the use of new technologies and data, are not covered by the Guidelines, and should be considered separately.

In addition, even if the reliability of ML components has been confirmed, when changes are made to facilities, systems, procedures, etc. as a result of its introduction, change management must be conducted, but the methods are not covered by the Guidelines.

The Guidelines assess the reliability of ML components, assuming that they are implemented in a specific plant[12]. When the same ML component is applied to multiple implementations, a separate assessment needs to be performed for each implementation.

In the Guidelines, reliability is assessed for a single ML component. In the case of a system with multiple ML components, each ML component needs to be assessed separately[13].

The Guidelines are not intended to relax or interpret the provisions of laws and regulations; one must comply with statutory obligations when using ML components for statutory inspections.

The responsibility of explaining to the regulatory authorities whether ML compoment development process is compliant to the Guidelines lies in the plant owner companies, should such explanation be necessary.

1.4 Structure and Reading of the Guidelines

The Guidelines are organized as follows.

Chapter 2, "Structure of Reliability Assessment of Machine Learning in the Field of Plant Safety," presents the hierarchical reliability assessment methodology of the Guidelines. The reader is required to construct and operate ML components in accordance with this chapter for appropriate reliability

---

[10] In the functional safety standards, it falls under the "JIS C 0508-3:2014 Functional safety of electrical/electronic/programmable electronic safety-related systems, Part 3: Software requirements."

[11] IEC 61511-1 (JIS C 0511-1) is the international standards for safety instrumented systems (functional safety) in the process industries. Here, software is classified as being realized in one of the following three languages: 3.2.75.1 FPL (fixed program language), 3.2.75.2 LVL (limited variability language), and 3.2.75.3 FVL (full variability language). ML components should be considered as elements that are implemented by FVL. Since FVL is described in IEC 61508-3:2010 (JIS C 0508-3:2014), the "reliability of ML components in the field of plant safety" covered by the Guidelines is considered to fall under IEC 61508-3:2010 (JIS C 0508-3:2014) as described in footnote 10.

[12] In the case of introducing a trained ML component, reliability of output cannot be assessed with existing methods, so the reliability is assessed with the method of the Guidelines assuming a specific implementing plant.

[13] The Guidelines take the position that it is difficult at this point to ensure redundancy and improve reliability with multiple ML components. For example, an ML component that aims to improve accuracy in an ensemble is assessed as a single ML component in the Guidelines.[14]

https://www.fdma.go.jp/relocation/neuter/topics/fieldList4_16/jisyuhoan_shiryo.html

assessment.
In Section 2.1, "Three Qualities of Reliability Assessment," the meaning of each of the three levels of reliability assessment ("quality in use," "external quality," and "internal quality") is presented, including their interpretation in the field of plant safety.
Section 2.2, "Methods and Requirements of Reliability Assessment," presents methods and criteria for conducting reliability assessment based on the three levels of quality.
Chapter 3, "Use Cases of Machine Learning in the Field of Plant Safety," presents a specific example of reliability assessment based on typical use cases in the field of plant safety. This chapter is intended to be referred to by the reader when practicing the contents of Chapter 2, and is expected to be applied to ML components that the reader builds and operates.
In Chapter 4, "Flow of Using the Guidelines," the implementation items of a reliability assessment are organized according to the flow of ML based system development and operation, showing examples of how to use the Guidelines in each phase and each entity involved. This chapter is intended to be referred to by the reader when practicing the contents of Chapter 2, and is expected to be used flexibly according to the circumstances of the reader's project.
In addition, as an appendix, the requirements for internal quality among the three levels of quality and the points to keep in mind for executing the requirements in the field of plant safety are organized and presented as a checklist.
In order to ensure that the implementations described in the Guidelines are achieved without fail, a template recording the results of reliability assessment process is also available[14]. By recording the results of reliability assessment according to the template, evidence of consulting and applying the Guidelines can be kept and used for explanations within and without the company. In addition, records of actual reliability assessments using the template are available as “Practical Examples” for all use cases. By referring to Practical Examples that are similar to one’s own case, it is possible to consider the implementation of reliability assessment and the level of detail in the template while referring to specific examples.

The Guidelines are organized in such manner that readers can understand the methods, applications, and implementation items of reliability assessment by reading them in order starting from Chapter 1. But it is also assumed that readers may start from specific chapters and sections according to their interests. The following table shows the relevant section in the Guidelines according to the matter of interest.

[14] https://www.fdma.go.jp/relocation/neuter/topics/fieldList4_16/jisyuhoan_shiryo.html

Table 1-2　Relevant section in the Guidelines according to the matter of interest

Item of Interest	Relevant Section in the Guidelines
Understanding which section of the Guidelines should be understood by whom (e.g. Quality Assurance staff, Plant System staff, and ML Development and Design staff)	Section 4.1
Understanding who will use the Guidelines and how	Section 4.2
Understanding examples of reliability assessment similar to the AI considered by the company	Check the positioning and list of use cases in Sections 3.1 and 3.2, and select a use case → Refer to the relevant use case in Section 3.3 See also "Practical examples" 3.3.1: Prediction of pipe wall thickness Practical Example: “1. Prediction of pipe wall thickness (Yokogawa Electric Corporation)” 3.3.2: Pipeline image diagnosis Practical Example: “1. Prediction of pipe wall thickness (Yokogawa Electric Corporation)” 3.3.3: Equipment deterioration diagnosis Practical Example: “1. Prediction of pipe wall thickness (Yokogawa Electric Corporation)” 3.3.4: Detection and diagnosis of early signs of abnormality Practical Example: “4-1. Prediction and diagnosis of abnormality (Chiyoda Corporation and Seibu Oil Company Limited)” Practical Example: “4-2. Prediction and diagnosis of abnormality (JGC Japan Corporation)” 3.3.5: Optimization of operation Practical Example “5-1. Optimization of operation (ENEOS Corporation and Preferred Networks, Inc.)”” Practical Example: “5-2. Optimization of operation (Yokogawa Electric Corporation,

	JSR Corporation)”
Understanding the meaning of three qualities of reliability assessment (quality in use, external quality, and internal quality)	Section 2.1
Understanding the specific method of reliability assessment based on the three qualities of reliability assessment	Section 2.2
Checking the specific requirement and perspectives of internal quality	Appendix
Confirming the definition of terms	Section 1.5
Recording the application status of the Guidelines	Template for Reliability Assessment Records

1.5 Terminologies

The terminologies used in the Guidelines shall be defined as follows.

(1) Machine learning

An artificial intelligence system, especially one in which a computer system automatically recognizes patterns in data and uses the results to make inferences and decisions without explicit program instructions. Deep learning is also a typical example of machine learning. [15]

(2) Machine learning (ML) component

A software component that is implemented by applying machine learning technology. [13] In the field of plant safety, for example, it refers to software that detects signs of abnormality from process data, and software that determines the degree of corrosion from images of pipes.
The Guidelines only cover the quality of ML components.

(3) Machine learning (ML) based system

A system that contains ML elements as the components. [16]In the field of plant safety, it refers to, for example, detection and diagnosis systems of early signs of abnormality that utilize machine learning, control systems that incorporate optimization of operation functions based on machine learning, and systems that determine and visualize the degree of corrosion from images of pipes taken by drones.
The Guidelines do not cover the quality of elements other than the ML components (software/hardware) that are contained in the ML based system.

(4) Supervised Learning/Unsupervised Learning/Reinforcement Learning

There are three types of learning methods in machine learning: supervised learning, unsupervised learning, and reinforcement learning.
“Supervised learning” is a learning method that predicts and identifies patterns in the output data from the input data. It learns the relationship between input and output by using a given set of input data and a set of correct output data as training data. Typical examples of problems solved by supervised learning are "regression" and "classification." In the Guidelines, the use cases of "prediction of pipe wall thickness (regression) (3.3.1),” "pipeline image diagnosis (classification) (3.3.2),” and "equipment deterioration diagnosis (classification) (3.3.3)” are applicable.
“Unsupervised learning” is a learning method in which the correct answer is not given as training data. A typical example is "clustering," which groups training data according to their features. In the Guidelines, the use case of "detection and diagnosis of early signs of abnormality (3.3.4)" applies.
Reinforcement learning is a method of learning to maximize the reward in a framework in which a reward is obtained by choosing an action in a certain environment. Training data do not contain correct answers. In the Guidelines, the use case of "optimization of operation (3.3.5)" applies.

---

[15] “AIST Announces Guidelines on Quality Control for Machine Learning," https://www.aist.go.jp/aist_j/press_release/pr2020/pr20200630_2/pr20200630_2.html, viewed on September 8, 2020

[16] National Institute of Advanced Industrial Science and Technology (AIST) (2020), “1st edition of Guidelines on Quality Control for Machine Learning”

(5) Regression model/Classification model

In supervised learning, the problem of predicting continuous values such as pipe wall thickness is called a regression problem, and the problem of discriminating discrete values (categories) such as the presence or absence of corrosion in pipes is called a classification problem. The models used in each problem are called a regression model and a classification model. The Guidelines provide use cases of "prediction of pipe wall thickness (3.3.1)" as a regression model, and "pipeline image diagnosis (3.3.2)" and "equipment deterioration diagnosis (3.3.3)" as examples of a classification model.

(6) Plant

Refineries, chemical plants (including petrochemical plants) and other sites, including petroleum complex areas. [17]

(7) Safety-related systems

Systems that satisfy both of the following.

- Perform the required safety functions necessary to shift equipment under control (EUC) to a safe state or to maintain the safe state of EUC.
- Achieve the level of safety necessary for the required safety functions by themselves or through other E/E/PE (electrical/electronic/programmable electronic) safety-related systems and other risk mitigation measures. [18]

Under the Guidelines, EUC refers to equipment, machinery, devices, and plants used mainly for manufacturing and maintenance. [16] The term E/E/PE system refers to a system that includes elements such as power source supply, input devices (sensors), interfaces and other communication paths, and output devices (actuators, etc.) [16]

In the Guidelines, the term "safety-related systems" assumes that ML components are included in safety-related systems. If the impact of misjudgment of ML components is greater than a certain level, SIL assessment of the entire safety-related systems is required. When referring to safety-related systems that do not include ML components, the following paragraph (8) "Safety-related systems independent of the ML based system" is used.

(8) Safety-related systems independent of the ML based system

Safety-related systems ensure the safe state of a target piece of equipment regardless of the input, processing, and output of ML based systemML based system. They are different from "external safety mechanisms" that compensate for undesired outputs of ML components.

The Guidelines also state that "risk avoidance" to ML components may not be necessary when Safety-related systems independent of the ML based system are present (Section 2.3.2).

(9) Safety Integrity Level (SIL)

A discrete level corresponding to a range of safety values as defined in the functional safety standards

---

[17] the Liaison Council of Three Ministries on Disaster Prevention of Petroleum Complexes (2019), "Guidelines for the Safe Operation of Drones in Plants"

[18] JIS C 0508-4: 2012 Functional safety of electrical/electronic/programmable electronic safety-related systems, Part 4: Definitions and abbreviations of terms

IEC 61508 (JIS C 0508) and IEC 61511 (JIS C 0511). Safety Integrity Level 4 is the highest safety level, and 1 is the lowest. [19]

Safety integrity refers to the probability[17] that E/E/PE safety-related systems will perform a stipulated safety function under all designated conditions within a specified period. SIL is used to stipulate safety requirements for safety functions.

The Guidelines do not allow ML components to be assigned SIL 2 or higher out of the four SILs (1 to 4), and set the level of "risk avoidance" for ML components within the range of "SIL 1" or "no SIL."

(10) External safety mechanism

It is software or hardware that is processed in parallel or in series with ML components for the purpose of improving safety, that monitors and corrects undesired output of ML components (limits or overwrites the output), and that can be assessed to be sufficiently safe using existing system development process methods that follow functional safety standards such as IEC 61508 (JIS C 0508) and IEC 61511 (JIC C 0511), particularly for software, IEC 61508-3 (JIS C 0508-3), etc.[20]In the Guidelines, the required level of "risk avoidance" changes depending on whether or not there is an external safety mechanism, i.e. if there is an external safety mechanism, the required level of "risk avoidance" decreases.

Figure 1-2 shows the relationship between (2) ML components, (8) Safety-related systems independent of the ML based system, and (10) external safety mechanism.

The Guidelines classify the human involvement from the output of ML components to the decision to act on the plant facilities into three categories (described in detail in 2.2.3 (1) 2) b. See Figure 2-7.). "External safety mechanism" is, regardless of the method of human involvement, positioned to override or correct the output of ML components based on logic, i.e. if the output of an ML component exceeds a certain range or is inconsistent with the output or judgment of another system, the output or judgment of the ML component is stopped.

“Safety-related systems independent of the ML based system (interlock, etc.)" are safety-related systems that ensure the safe state of the target equipment regardless of the input, processing, and output of the ML based system, and are different from an "external safety mechanism.”

[19] JIS C 0508-4: 2012 Functional safety of electrical/electronic/programmable electronic safety-related systems, Part 4: Definitions and abbreviations of terms

[20] The Guidelines define matters with reference to the “Guidelines on Quality Control for Machine Learning,” the National Institute of Advanced Industrial Science and Technology (AIST), 2020.

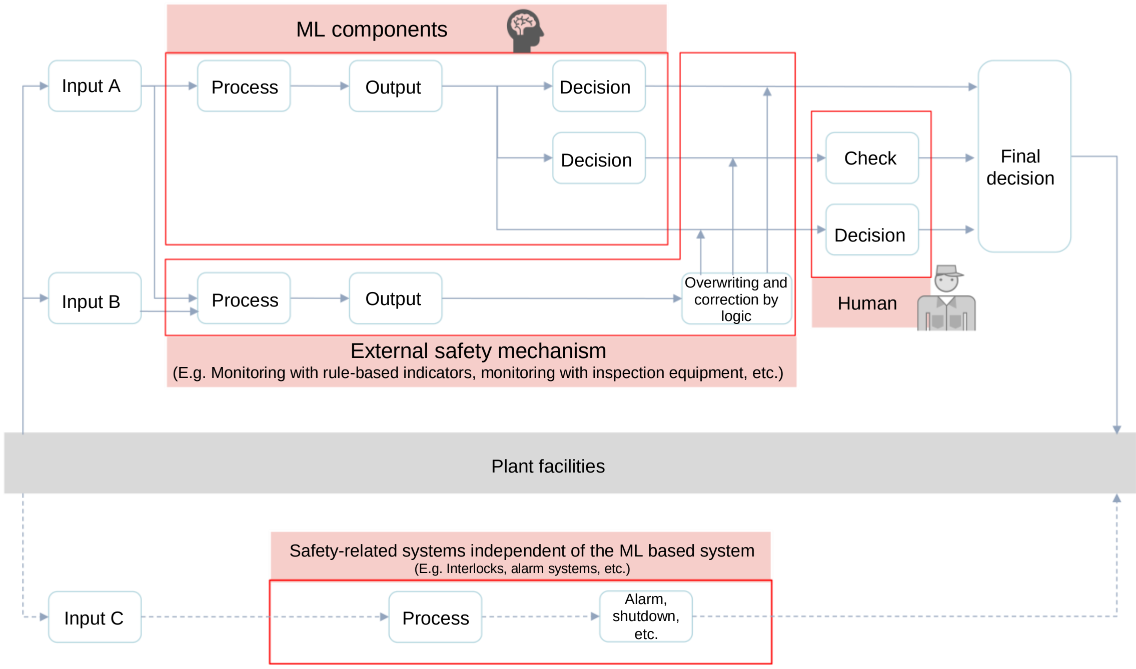


Figure 1-2 Relationship between ML components, external safety mechanism, and Safety-related systems independent of the ML based system

(11) Reliability (software reliability)

The ability to maintain a certain performance when used under specified conditions. [21]
In the Guidelines, it refers to the ability of ML components to perform as expected, e.g. the reliability of ML components to determine corrosion from images of piping.
Since the Guidelines describe the reliability of ML components, the term "reliability" used alone refers to software reliability.

(12) Risk

Probability of harm and a combination of the degree of the harm. [22]
The only risk addressed in the Guidelines is the risk associated with misjudgment of ML components.

(13) Safety

The absence of unacceptable risks. [23]
The Guidelines define safety as a state in which the risk associated with misjudgment of ML components is controlled at an acceptable level.

---

[21] JIS Z 8115: 2019 Dependability (overall reliability) terminologies
[22] JIS Z 8051: 2015 Safety Aspects - Guidelines for introduction to the standards
[23] JIS C 0511-1: 2019 Functional safety - Safety instrumented systems for the process industries - Part 1: Framework, definitions, system, hardware and application programming requirements

(14) Minor injury

Of accidents without lost workdays, those at a scratch level that do not require medical attention. (So-called "non-serious injuries")
One of the severity standards for determining the impact of misjudgment of ML components when setting the demand level of risk avoidance.

(15) Quality in use

The quality of the system as a whole that should be provided to the end-users[24] (how well it meets the users' needs (performance, usability, etc.) in a particular usage situation). In the field of plant safety, "end-users" may be an operator/maintenance engineer, a manager responsible for the safety and productivity of individual pieces of equipment, the head of the plant, or the management of the plant owner company, depending on the purpose of the ML based system.
The Guidelines describe the realization of quality in use items that correspond to the external quality of ML components, not quality in use items that correspond to non-ML components.

(16) External quality

A quality expected to be satisfied by the components of the system that are constructed by machine learning. [22]
By achieving external quality, quality in use is achieved.
In the Guidelines, the term "external quality" used alone refers to the external quality of ML components.

(17) Risk avoidance (one of the two axes of quality in use and external quality)

It is a quality characteristic that avoids undesirable decision-making behavior of ML components[25], which may cause adverse effects, such as human suffering or economic loss, to the operators and users of the product, or third parties. Establishing and realizing the requirements for "risk avoidance" corresponds to the concept of "risk reduction" in the field of safety. [22] It is one of the two types of quality in use and external quality.
The Guidelines refer to the avoidance of adverse effects on plant safety, such as oversight of abnormality in detection and diagnosis, and corrosion in pipeline image diagnosis. See Section 2.1.3 for details.
The term "avoidance" here does not mean "risk avoidance (withdrawal from activities)," which is one of the risk responses in risk management[26].

(18) Performance (one of the two axes of quality in use and external quality)

An axis that collectively refers to the quality in use and external quality seeking a decision that ML components are conducive to enhancing productivity and efficiency. Specifically, it refers to the output

---

[24] National Institute of Advanced Industrial Science and Technology (AIST), "1st edition of Guidelines on Quality Control for Machine Learning"

[25] Risk Avoidance in the Guidelines aims to avoid adverse effects caused by misjudgment of ML components. The Guidelines do not cover the adverse effects accompanying changes in facilities, systems, procedures, etc. when introduced ML components are confirmed to be sufficiently reliable. Even if the reliability of ML components is confirmed, it is necessary to perform change management so as not to cause adverse effects associated with changes.

[26] JIS Q 0073: 2010 Risk management - Terminologies

expected by users of the ML based system (excluding the output to pursue "assurance and safety" included in "risk avoidance") with higher accuracy and probability on average over the long term. In outputs where adverse effects of individual misjudgments are not a major issue, seeking high average performance is more essential than the necessity of individual outputs. [27]
In the Guidelines, it implies an expectation for demonstrating higher performance on average over the long term, while the frequency of misjudgments in detection and diagnosis of abnormality (alerting abnormality when there is no abnormality) and safety errors in prediction of pipe wall thickness (judging treatment to be necessary when no such treatment is required), and other individual misjudgments are acceptable. See Section 2.1.3 for details.

(19) Internal quality

The inherent quality of a component of machine learning. [28]It is what must be satisfied in the design, development, and operation of ML components in order to satisfy "external quality," and there are eight axes for the following three types of quality (each quality is detailed in Section 2.1.4).

- Appropriateness of data used in development (e.g. amount and type of data)
- Appropriateness of the model developed (e.g. accuracy of machine learning during testing)
- Appropriateness of implementation and operation methods (e.g. how to maintain the accuracy of machine learning)

(20) Use case

A typical example of the use of an ML based system in the field of plant safety. The Guidelines provide examples of quality in use and external quality items based on five use cases, as well as points to keep in mind when executing internal quality requirements. See Chapter 3 for details.

(21) Perspective (in the field of plant safety/unique to use case)

Points to keep in mind for executing the internal quality requirements described in the "Machine Learning Quality Management Guideline 1st edition" of the National Institute of Advanced Industrial Science and Technology (AIST) in the field of plant safety. There are "perspectives in the field of plant safety" that should be referred to commonly in the field of plant safety regardless of the use case, and "perspectives unique to the use case" that should be referred to specifically for the development of individual use cases. For example, there are following perspectives.

- In order to achieve the requirement of "unbiased data sampling," the use of simulators should be considered when the amount of measured data is biased (perspectives in the field of plant safety)
- In order to achieve the requirement of "unbiased data sampling" for the use case of detection and diagnosis of early signs of abnormality, it is not necessary to cover all data during non-normal operations, but rather to seek comprehensive sample extraction in the normal region (use case-specific perspectives).

(22) PoC

An abbreviation of "Proof of Concept." PoC refers to verification activities conducted to confirm the feasibility of a new idea prior to full-scale implementation or systemization.

---

[27] Description based on the "1st edition of Guidelines on Quality Control for Machine Learning," the National Institute of Advanced Industrial Science and Technology (AIST)
[28] "AIST Announces Guidelines on Quality Control for Machine Learning," https://www.aist.go.jp/aist_j/press_release/pr2020/pr20200630_2/pr20200630_2.html, viewed on September 11, 2020

PoC is a concept that includes a variety of meanings, from pure trial study to preparation for full-scale development. And although reliability assessment is not required in all cases, the Guidelines indicate items that should be confirmed in the PoC stage in Section 4.2.3.

(23) Review

Terminology used in Chapter 4 of the Guidelines to describe how to engage in the reliability assessment. Upon receiving a request from key personnel engaged in reliability assessments using the Guidelines, the items to be considered by key personnel are confirmed based on their own responsibilities and expertise.
A reviewer does not necessarily need to read and understand the contents of the guidelines, but should be involved in the reliability assessment in response to a request from the key personnel.

1.6 Reference: "Collection of Case Examples of Leading Companies Introducing AI into Plants"

As shown in Figure 1-3, there are various challenges to introducing AI in the field of plant safety. These include not only ensuring and assessing reliability, but also human resources and structures, and uncertainty of economic benefits. For an overall picture of these challenges, examples of overcoming challenges, and concrete results, please refer to "Collection of Case Examples of Leading Companies Introducing AI into Plants - Practical Examples of Achieving Results and Breaking Through Challenges in AI Projects -."

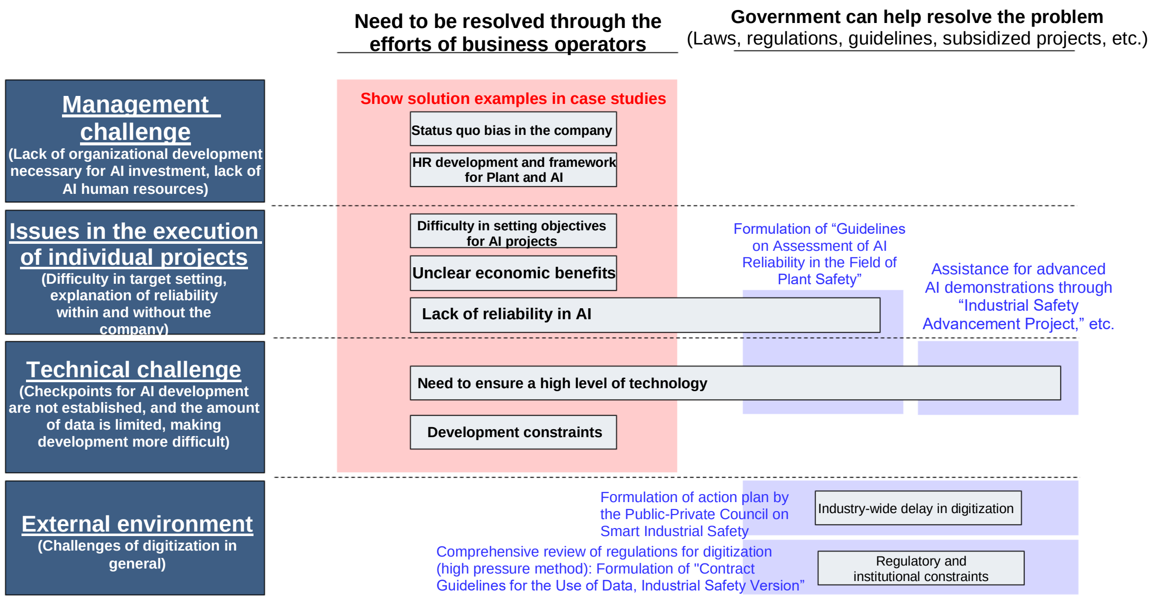


Figure 1-3 Classification of AI implementation issues in the field of plant safety and the positioning of "Collection of Case Examples of Leading Companies Introducing AI into Plants"[29]

[29] the Liaison Council of Three Ministries on Disaster Prevention of Petroleum Complexes (Ministry of Economy, Trade and Industry; Fire and Disaster Management Agency; and Ministry of Health, Labour and Welfare) (2020) "Collection of Case Examples of Leading Companies Introducing AI into Plants - Practical Examples of Achieving Results and Breaking Through Challenges in AI Projects -"

2. Reliability Assessment Structure of Machine Learning in the Field of Plant Safety

In the Guidelines, "the required level of 'external quality' is achieved by improving the 'internal quality' of ML components to realize the 'quality in use' of the final system"[30] using the hierarchical quality assurance structure of the "Machine Learning Quality Management Guideline 1st edition".
A schematic of the hierarchical quality assurance procedure is shown in Figure 2-1. (1) Define what is required to be achieved by the ML based system (quality in use), (2) Identify the external quality required of ML components to satisfy quality in use, and set the required level (i.e. the level of demand), and (3) Create ML components (internal quality) based on the requirements according to the level.
After explaining "quality in use," "external quality" and "internal quality" in 2.1, the abovementioned hierarchical reliability assessment method is detailed in 2.2.

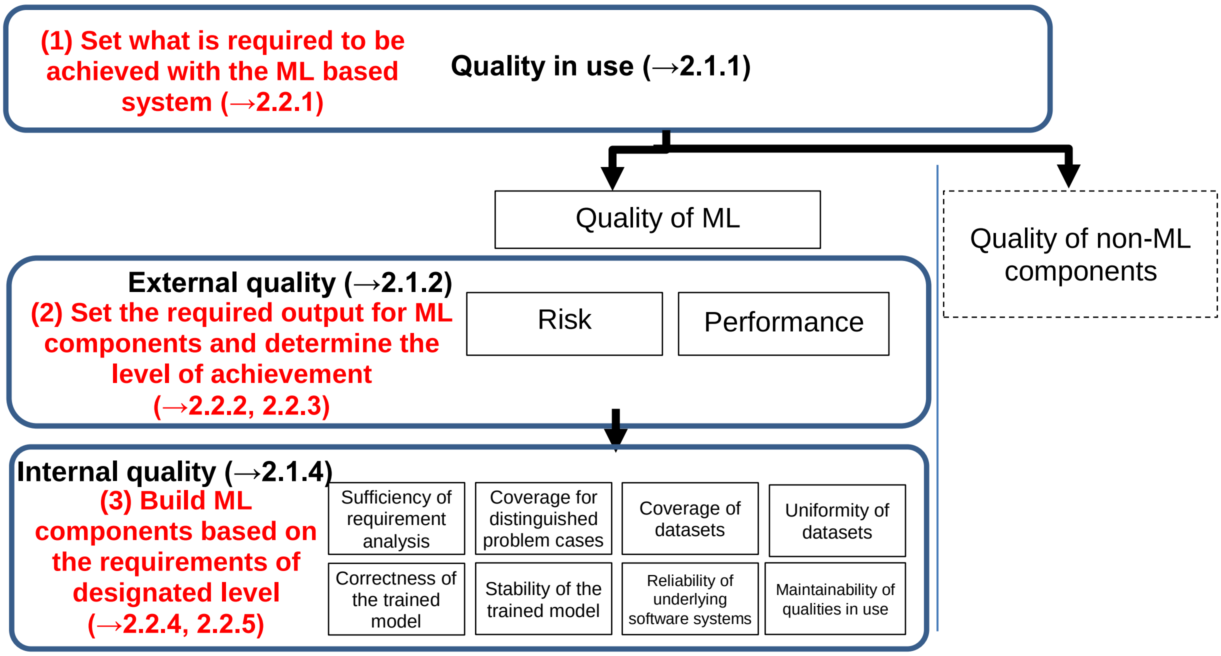


Figure 2-1 Hierarchical quality assurance of ML based system

Source: Prepared by Mitsubishi Research Institute based on the "Machine Learning Quality Management Guideline 1st edition", the National Institute of Advanced Industrial Science and Technology (AIST)

While the Guidelines only consider the quality of ML components, it is also important to ensure the quality of non-ML components (e.g. components comprising rule-based systems) in order to secure the quality of an ML based system. The quality of non-ML components should be secured according to existing quality assurance and assessment systems (e.g. international standards for functional safety and software quality). Furthermore, the quality of the ML based system depends on how the ML components are combined with other components (e.g. how to determine the rules for comparing the output results of the ML components with those of other components, requiring human judgment in case of inconsistency, and automatic operation in case of consistency). It is essential to consider not only the quality of the ML components targeted in the Guidelines, but also how machine learning can be used in the system to improve the quality of the entire ML based system.

[30] National Institute of Advanced Industrial Science and Technology (AIST), "1st edition of Guidelines on Quality Control for Machine Learning"

2.1 Three Qualities of Reliability Assessment

2.1.1 Quality in use

A quality[31] that the entire ML based system should provide to the end-users is the quality in use. In order to achieve this, the system components including ML components are developed.
The quality in use includes the objectives that the system users expect from the system and the safety that should be ensured as a premise; in other words, "what is required to be achieved by the system." In the field of plant safety, quality in use refers to the objectives expected by users and the safety that should be provided as a prerequisite when ML based system such as "detection and diagnosis system of early signs of abnormality" and "pipeline image diagnosis system" are introduced to a specific plant[32].

[Examples of quality in use]

- (In the case of detection and diagnosis system of early signs of abnormality)
  Correctly detect the occurrence of future abnormalities under various conditions
- (Pipeline image diagnosis system)
  Do not overlook piping that requires visual inspection

“System user" here may be an operator/maintenance engineer, a manager responsible for the safety and productivity of individual pieces of equipment, the head of the plant, or the management of the plant owner company, depending on the purpose of the ML based system, and the quality in use is specified from the standpoint of the system user concerned. For this reason, when quality in use is verbalized, it is mainly expressed qualitatively. In order to achieve quality in use as a system, it is necessary for the system components to perform at a prescribed level. This is the "external quality" in the next section.

2.1.2 External quality

Quality[28]expected to be satisfied by the system components using machine learning is the external quality.
The external quality is expressed as the performance required of ML components to achieve quality in use. In the field of plant safety, external quality refers to the performance required of ML components (detection of signs and determination of corrosion) in “detection and diagnosis system of early signs of abnormality” and “pipeline image diagnosis system.” [33]

Example of external quality

- (In the case of detection and diagnosis system of early signs of abnormality)

[31] National Institute of Advanced Industrial Science and Technology (AIST), “1st edition of Guidelines on Quality Control for Machine Learning”

[32] The “1st edition of Guidelines on Quality Control for Machine Learning” lists the following as specific examples of quality in use in other fields.
Example of automated vehicles: Safety against collision with obstacles under all drivable circumstances
Example of automatic stock transaction service: Maximization of profits

[33] The “1st edition of Guidelines on Quality Control for Machine Learning” lists the following as specific examples of external quality in other fields.
Example of an object recognition module installed in an automated vehicle: Correctly recognizing obstacles in all possible weather conditions and time zones
Example of a stock price prediction module included in the automated stock trading service: Minimizing an error in stock price prediction and maximizing the sum of expected trading results.

In the case of "signs of abnormality," reduce the false-negative rate as much as possible

- (Pipeline image diagnosis system)
  When a visual inspection is "required," reduce the false-negative rate as much as possible where it is judged to be "not required"

Considering that an ML based system consists of ML components and other components, quality in use is achieved by "external quality of ML components" and "external quality of other components." The "external quality of other components" is assumed to be ensured in accordance with existing quality assurance and assessment systems (such as international standards for functional safety and software quality), and in the Guidelines, the term "external quality" used alone refers to the “external quality of ML components.”

2.1.3 Axes of quality in use and external quality

The quality in use and external quality are classified according to their characteristics, which the Guidelines refer to as "axis." An "axis" is a classification of the quality to be achieved for individual ML based system and ML components. The Guidelines define two axes, "risk avoidance" and "performance," and they are described below[34]. All types of quality in use and external quality belong to one of the axes.
Regarding external quality, levels are defined for each axis of "risk avoidance" and "performance" according to the level of demand (see 2.2.3), and there is a structure in place, where the level of requirement for “internal quality” in the next section is decided according to the level of external quality.

(1) Risk avoidance

The "risk avoidance" axis is a collective term for the quality in use and external quality that avoids[35] the adverse effects of misjudgments of ML components[36] on “safety assurance,” such as human casualties and economic losses, to the operators and users of the ML based system and third parties. Establishing and realizing the requirements for the external quality that affiliate to the "risk avoidance" axis is equivalent to "risk mitigation" in the field of safety. [37]
In the field of plant safety, the following are examples of quality in use and external quality corresponding to "risk avoidance"[38].

[34] In the “1st edition of Guidelines on Quality Control for Machine Learning,” the three axes of external quality of ML components are “risk avoidance,” “performance,” and “fairness.” “Fairness" is an axis that requires ML components to have social norms and ethics in machine learning, because ML based systems affect consumers and other citizens. However, in the field of plant safety, the target of machine learning is plant equipment, not citizens, so it was judged that the axis of "fairness" is not necessary in the Guidelines. Although not covered in the Guidelines, the axis of "fairness" may need to be considered for use cases that include data such as images, video data, and conversations of employees.

[35] The term “avoidance” here does not mean "risk avoidance (withdrawal from activities)," which is one of the risk responses in risk management.

[36] Risk Avoidance in the Guidelines aims to avoid adverse effects caused by misjudgment of ML components. The Guidelines do not cover the adverse effects accompanying changes in facilities, systems, procedures, etc. when introduced ML components are confirmed to be sufficiently reliable. Even if the reliability of ML components is confirmed, it is necessary to perform change management so as not to cause adverse effects associated with changes.

[37] National Institute of Advanced Industrial Science and Technology (AIST), “1st edition of Guidelines on Quality Control for Machine Learning”

[38] Section 3provides examples of external quality items for five use cases.

- (In the case of detection and diagnosis system of early signs of abnormality)
  Quality in use: Correctly detect the occurrence of future abnormalities under a variety of plant conditions
  External quality: In the case of "signs of abnormality," reduce the false-negative rate as much as possible where it is judged to be "normal"
  (Quality to avoid cases where ML components miss signs of abnormality and cases where abnormality actually occurs)
- (In the case of pipeline image diagnosis system)
  Quality in use: Do not overlook piping that requires visual inspection
  External quality: When a visual inspection is "required," reduce the false-negative rate as much as possible where it is judged to be "not required"
  (Quality to avoid serious damage or accidents from occurring as a result of ML components determining that degraded piping is safe)

This refers to the quality of avoiding adverse effects on the safety of the plant, such as an oversight of abnormality and corrosion in pipeline image diagnosis. For such items, the level of external quality shall be set according to the level of demand, etc. to avoid adverse effects. (See 2.2.3 (1) for details of how to set specific levels.)

(2) Performance

An axis that collectively refers to the quality in use and external quality requiring ML components to make decisions related to productivity and efficiency enhancement. In detail, the term refers to the quality of producing outputs expected by ML system users with higher accuracy and probability on average in the long term (excluding outputs pursuing “safety assurance”, a part of “risk avoidance”). In outputs where adverse effects of individual misjudgments are not a major issue, seeking high average performance is more essential than the necessity of individual outputs. [39]
In the field of plant safety, the following are examples of quality in use and external quality under the "performance."

- (In the case of detection and diagnosis system of early signs of abnormality)
  Quality in use: Set the alarm frequency to a reasonable level so that the operators and inspectors do not have to allocate extensive time resource to check the contents of the alarm
  External quality: Reduce the frequency of false positives below a certain level
- (In the case of pipeline image diagnosis system)
  Quality in use: Reduce the number of visual inspections conducted by maintenance engineers
  External quality: Reduce the false-positive rate to a certain level, so that a visual inspection is not diagnosed as “required” where in fact it is “not required.”

This item refers to the quality where long-term performance in average (i.e. False-positives and miscalculations of hazard are minimized) is expected, although individual misjudgment e.g. frequency of false-positive (where an alert is made when there is no abnormality) and miscalculation of hazard (where an action is deemed necessary when there is no need) may be acceptable. The level of external quality shall be set so that these demands can be satisfied (See 2.2.3 (2) for details on how to set the levels).

---

[39] Description based on the “1st edition of Guidelines on Quality Control for Machine Learning,” the National Institute of Advanced Industrial Science and Technology (AIST)

It should be noted that "risk avoidance" and "performance" are not necessarily contradictory, and they are often required simultaneously on the same ML based system. For example, in the case of a system used for maintenance, "risk avoidance" is required to avoid oversight of deterioration, and "performance" is also required to minimize the frequency of inspection and replacement. It is required to set external quality levels for each axis, and construct internal quality based on the highest level.

2.1.4 Internal quality

Since ML components do not judge abnormal/normal, etc. through deductive programming by developers, misjudgments can be caused by various factors. These include insufficient training dataset, overfitting on training dataset, and insufficient adaptation to changes in the implementation environment. Therefore, in order to control external quality, it is not enough to validate a program (codes) — it is necessary to comprehensively validate the entire process of machine learning, from design to operation. For this purpose, the Guidelines set two internal qualities for each of the following categories: (1) data quality design, (2) data quality, (3) model quality, and (4) quality of implementation and operation, and seek realization of external quality through their management. These internal qualities are common to all ML components, and unlike quality in use and external quality, readers do not need to set them by themselves (See 2.2.1 for setting the quality in use and 2.2.2 for setting the external quality).

In the ”Machine Learning Quality Management Guideline 1st edition”, the items to be implemented to ensure each type of internal quality are summarized as requirements, and are classified into three levels according to the level of external quality. ML components that satisfy the external quality are implemented by building ML components according to the requirements that apply to the required level (the satisfaction of external quality is confirmed by testing). Checking the level of internal quality is detailed in 2.2.4, and checking and execution of internal quality requirements are detailed in 2.2.5.

In the Guidelines, the requirements in the ”Machine Learning Quality Management Guideline 1st edition” are used without modification. As a supplement, the "perspectives" that need to be followed when applying the cross-sectoral requirements to the field of plant safety are summarized in "Appendix: “Perspectives in the Field of Plant Safety” Checklist for Internal Quality Assurance." Reflected in this section are the considerations listed in Chapter 7 "Industrial Process" domain of "Guideline for quality assurance of AI-based products 2020.08" (detailed in 1.2.1 (2)). The typical "perspectives" unique to the use cases in the field of plant safety are also described in the abovementioned Appendix (See Chapter 3 for details of use cases).

In the following, the eight axes of internal quality in the ”Machine Learning Quality Management Guideline 1st edition”, which are summarized in Figure 2-1, are explained for each of the following classifications: "data quality design," "data quality," "model quality," and "quality of implementation and operation.” Note that only examples of the requirements are given here, and the whole is summarized in the Appendix.

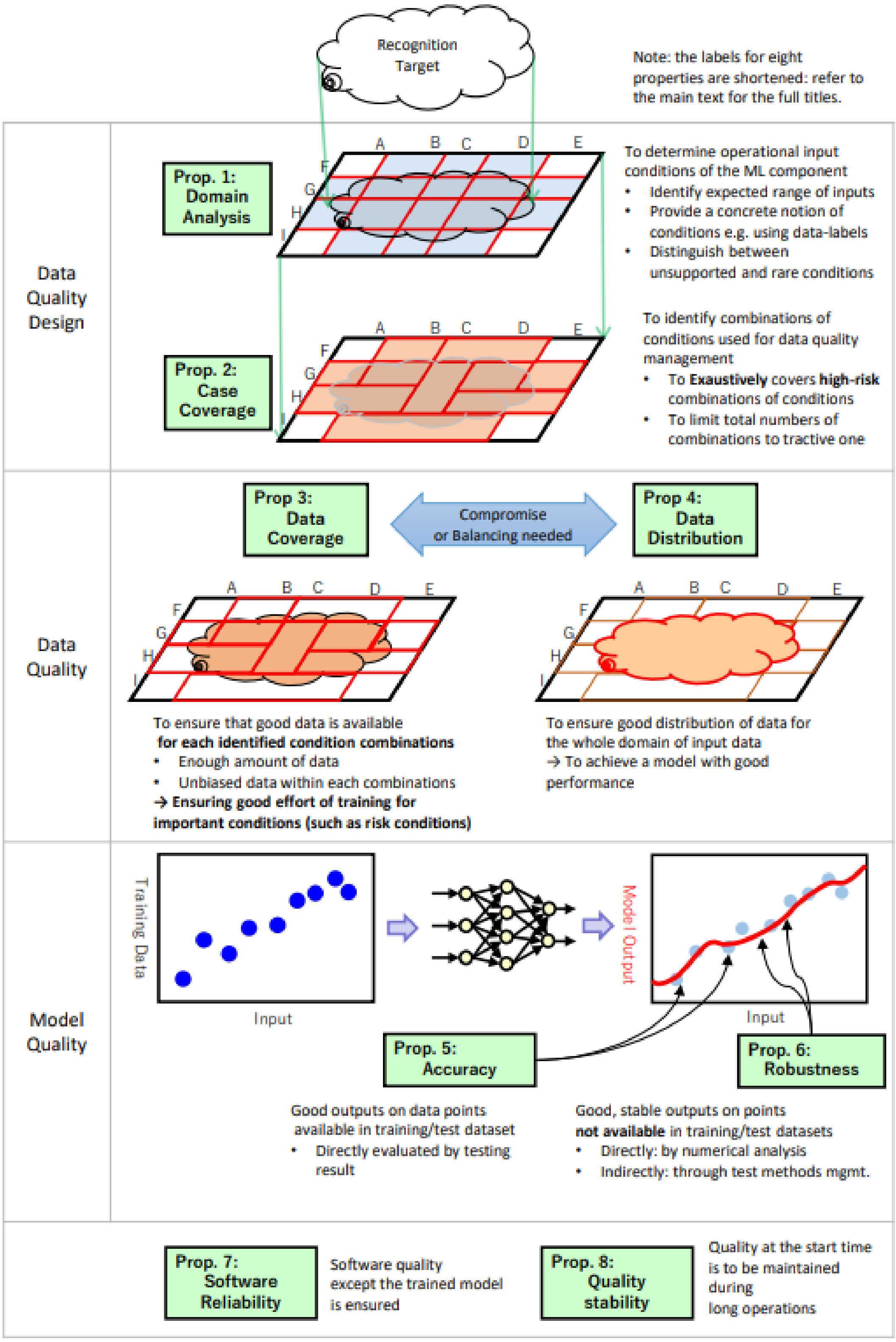


Figure 2-2 Eight axes of internal quality and their relationships

Source: National Institute of Advanced Industrial Science and Technology (AIST) "Machine Learning Quality Management Guideline 1st edition"

Axis related to Dataset quality design

(1) Sufficiency of requirement analysis

<u>Definition in the ”Machine Learning Quality Management Guideline 1st edition”</u> The characteristics of actual data during operation that is to be input into ML components corresponding to the real-world usage of the ML based system is analyzed, and the analysis results cover all assumed usage situations.
<u>Examples of requirements in the ”Machine Learning Quality Management Guideline 1st edition”</u> ✓ Specify the target and range of operations that ML components should address ✓ Identify the range of input data for ML components ✓ Determine the situations that ML components do not address, or that occur infrequently ✓ Consider the risk of quality degradation of ML based system caused by ML components

(Guidance)
Define the range of operations addressed by ML components. Identify the range of input data to be assumed, describe it in a concrete form such as data labels, and clearly distinguish situations that are not addressed, rare, etc.

(2) Coverage for distinguished problem cases

<u>Definition in the ”Machine Learning Quality Management Guideline 1st edition”</u> Premising the sufficiency of requirement analysis, full consideration should be given to the dataset design in order to collect and organize sufficient training and testing data for the various situations to be handled by the system.
<u>Examples of requirements in the ”Machine Learning Quality Management Guideline 1st edition”</u> ✓ Cover cases where the quality of an ML based system is at risk of degradation ✓ Data attributes and data volume should be within a manageable range ✓ If the number of cases are limited, check them comprehensively; otherwise, check a sampled portion of the cases to cover all attributes and combinations ✓ When high quality is required, mathematical "exhaustivity criteria" should be introduced to the case extraction process

(Guidance)
Design a framework of dataset organization in subdivision to be used for quality control. This includes exhaustive coverage of combinations of high-risk situations, etc., as well as limiting the size of the dataset to a manageable level.

Axis related to Data Quality

(3) Coverage of datasets

<u>Definition in the "Machine Learning Quality Management Guideline 1st edition"</u> For each combination of situations to be addressed, there are no missing situations and a sufficient amount of training data is provided.
<u>Examples of requirements in the "Machine Learning Quality Management Guideline 1st edition"</u> ✓ The necessary data should be secured exhaustively by devising ways to construct the dataset ✓ If rare data cannot be obtained, consider solutions separately for each case via methods like testing. ✓ If the exhaustivity criteria are introduced, inspect whether attributes "not included in the case" are distributed without bias

(Guidance)
Make sure that each subdivided category contains sufficient data. Make sure that the amount of data is sufficient and unbiased. This will ensure that the training is sufficiently risk-responsive.

(4) Uniformity of datasets

<u>Definition in the "Machine Learning Quality Management Guideline 1st edition"</u> Each situation or case in the dataset should be extracted according to the frequency of occurrence in overall input data.
<u>Examples of requirements in the "Machine Learning Quality Management Guideline 1st edition"</u> ✓ The frequency of occurrence of individual attribute values should be appropriately monitored, while paying attention to avoid bias in the process of acquiring the entire data set ✓ Examine how to balance the uniformity and the coverage of the dataset, and design the dataset accordingly.

(Guidance)
Make sure that the occurrence of each situation or case is identical in frequency in both training dataset and overall input data. This is intended to improve the overall performance of the model.
"Coverage of datasets" and "Uniformity of datasets" need to be balanced.
Note that "Uniformity of datasets" has requirements according to the levels of "risk avoidance" and "performance."

Axes related to Model Quality

(5) Correctness of the trained model

<u>Definition in the "Machine Learning Quality Management Guideline 1st edition"</u> The ML components should respond as expected to the specific input data contained in the training dataset (consisting of training data, test data, and validation data).
<u>Examples of requirements in the "Machine Learning Quality Management Guideline 1st edition"</u> ✓ From the imput of the dataset, the ML components shall provide appropriate output that satisfys the requirements set by external quality. ✓ Evaluate the ML model by methods including changing data size and cross-validation. ✓ If a certain level of misjudgment is agreed as acceptable in the output, its judgment criteria should be defined

(Guidance)
Sufficiently accurate output is provided for the input from the dataset. The quality of training and test data is assessed directly based on the test results.

(6) Stability of the trained model

<u>Definition in the "Machine Learning Quality Management Guideline 1st edition"</u> For input data that are not part of a training dataset, ML components respond in a way that is sufficiently similar to the response to near-identical data in the training dataset.
<u>Examples of requirements in the "Machine Learning Quality Management Guideline 1st edition"</u> ✓ For input data that are not part of a data set, the output is similar to the output of a dataset close to the input data

(Guidance)
Sufficiently robust output is provided even from input data not in the training dataset. This should be ensured by numerical evaluation/analysis and testing methods. Ensuring the "stability of the trained model" is of particular importance, as in fields where safety is a necessity, maintaining stable performance against data in actual operation is vital.

Axis related to Quality of Implementation and Operation

(7) Reliability of underlying software systems

Definition in the "Machine Learning Quality Management Guideline 1st edition"
Training programs used in the training phase of machine learning, or the prediction and inference program used at runtime, must work correctly as a software program for the given data or trained machine learning model.
Examples of requirements ✓ Use reliable software ✓ Consider in advance the differences between the development and operation environments of ML components and their impact

(Guidance)

Quality of software for non-machine learning models needs to be ensured. Integrity of software, etc. used to develop ML components is required. It is assumed that non-ML components are developed in accordance with the quality and requirements expected of systems in general.

(8) Maintainability of qualities in use

Definition in the "Machine Learning Quality Management Guideline 1st edition"
The internal quality that was satisfied at the start of operation is maintained throughout the operation period.
Examples of requirements ✓ Define in advance the update frequency of ML components or the criteria to determine whether to update the components. ✓ Examine the method of quality test at the time of update, especially the judgment criteria (or decision-making method) for whether or not to update

(Guidance)

Quality at the start of operation needs to be maintained during the operation period. Requirements are items that should be considered in advance to maintain quality during operation. Particularly at the plants, changes take place over time due to various factors such as product switches and maintenance. It is important to ensure the "Maintainability of qualities in use" because changes in the state of equipment and acquisition of new data through the operation of an ML based system may lead to a failure to maintain the initial internal quality, "risk avoidance" and "performance."

2.2 Method of and Requirements for Reliability Assessment

In this section, the method of applying reliability assessment and requirements for ML components are described in detail according to the three levels of quality (quality in use, external quality, and internal quality). When conducting a reliability assessment, quality items should be specified, and levels should be set according to the procedures in this section. A diagram of the relationship between the three levels of quality used to explain the application method is shown below.

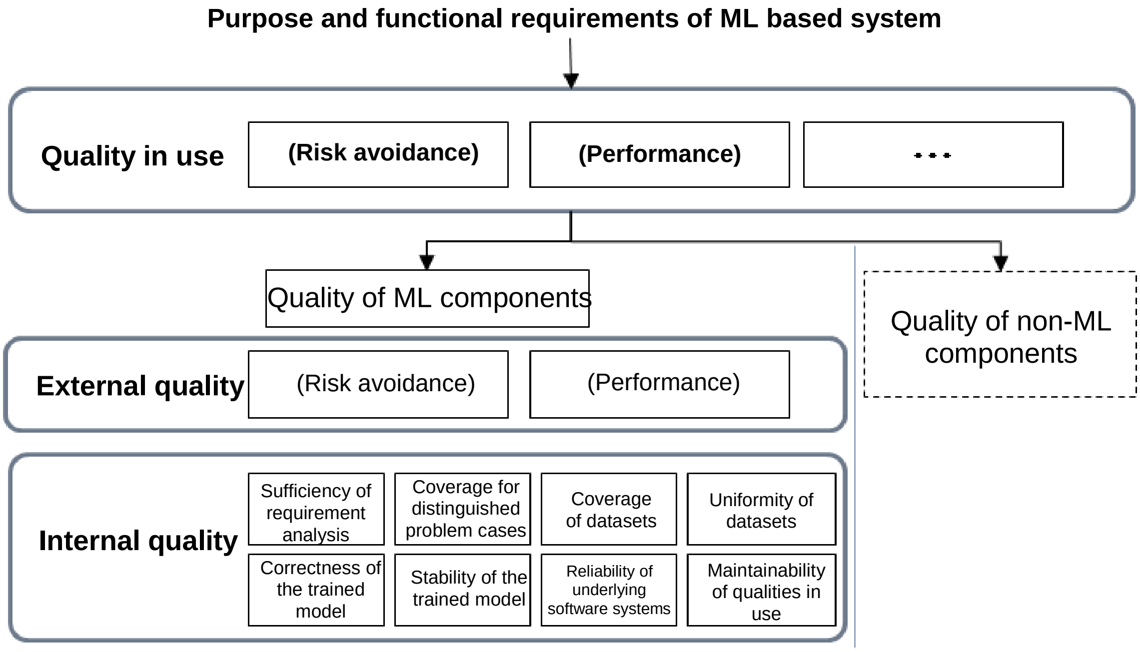


Figure 2-3 Relationship between three levels of quality in reliability assessment

2.2.1 Setting the quality in use

Based on the functional requirements of ML based system (what is requiredto be achieved), set the quality in use from the user perspective.

Figure 2-4 shows a conceptual image of setting quality in use based on the functional requirements of ML based system. [40]For example, in the case of an ML based system that performs "detection and diagnosis of early signs of abnormality," the plant operator checks the system output and performs the necessary operations for the plant. Therefore, quality in use is the quality required by the operator, who may be expected to "correctly detect the occurrence of future abnormalities under a variety of plant conditions" (applies to the axis of "risk avoidance") and "set a reasonable alert frequency that does not cause operators or inspectors to spend time on checking the contents of an alert" (applies to the axis of "performance").

[40] In Section 3, examples of quality in use items for each use case are described.

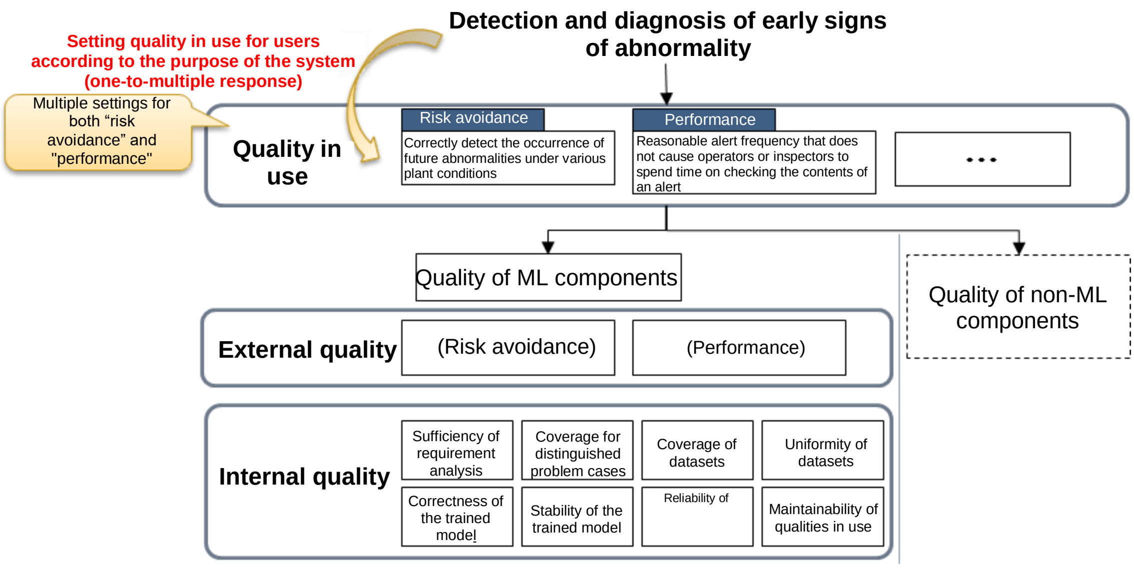


Figure 2-4　Configuration image of quality in use (example of detection and diagnosis system of early signs of abnormality)

2.2.2　Setting the external quality

The next step is to set the external quality of ML components, which corresponds to the quality in use set in the previous section. External quality is a quality required for the output of ML components and is usually set one-to-one with the quality in use. At the stage of setting the external quality in 2.2.2, a numerical target (e.g. accuracy of more than x%) specific to machine learning does not need to be set. As for the external quality of "risk avoidance," it may be possible to define numerical targets required for ML components (e.g. the incidence of misjudgments leading to danger) by checking safety-related systems and external safety mechanisms in the process of setting the level of external quality (2.2.3). Finally, at the stage of building ML components (2.2.5), specific numerical targets (e.g. Accuracy, F-measure) specific to machine learning can be set according to the results of PoC and the status of data acquisition and learning.

Figure 2-5 shows a conceptual image of setting the external quality for ML components from the quality in use of the ML based system. For example, in the case of an ML based system that performs “detection and diagnosis of early signs of abnormality,” the quality in use to "correctly detect the occurrence of future abnormalities under a variety of plant conditions" is corresponded by the external quality to "minimize the false-negative rate of judging ‘normal’ when ‘signs of abnormality’ are present" (corresponds to the axis of “risk avoidance”). In the same way, the quality in use to "set the alarm frequency to a reasonable level that does not require long hours for operators and inspectors to check the contents of the alarm" corresponds to the external quality to "minimize the number of misjudgments" (corresponds to the "performance" axis).

The quality in use consists of the external quality of "ML components" and "non-ML components," but the Guidelines only focus on the external quality of ML components.

In the field of plant safety, if Safety-related systems independent of the ML based system are established, consideration for “risk avoidance” may not be required for ML based system. For example, in the case of an ML based system that detects the degradation trend of equipment, if there is a separate sensor that detects the malfunction of equipment parts beyond a certain level and the safety aspect is guaranteed by the independent safety-related systems, the purpose of the ML based system is limited to the early detection of the degradation trend and the creation of an efficient maintenance plan. This eliminates the need to consider "risk avoidance" (see the use case in "3.3.3 Equipment deterioration diagnosis").

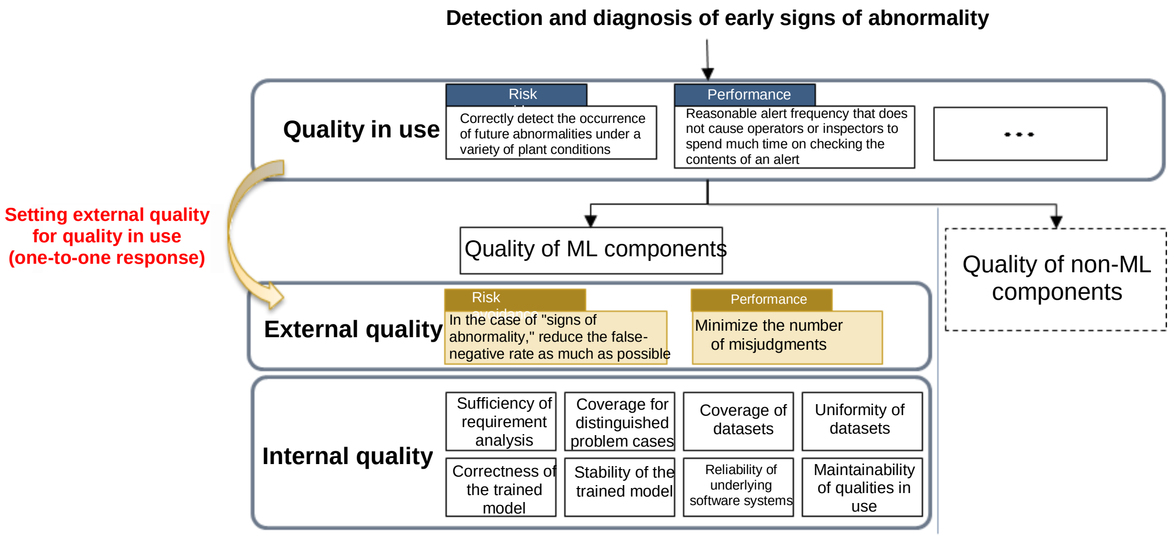


Figure 2-5 Conceptual image of setting the external quality (example of detection and diagnosis system of early signs of abnormality)

2.2.3 Setting the level of external quality[41]

For the set external quality, set the level according to the requirements for ML components. Using the definition of ”Machine Learning Quality Management Guideline 1st edition”, "risk avoidance" is set by AISL (AI Safety Level) and "Performance" is set by AIPL (AI Performance Level).

(1) Risk avoidance

The level of external quality for the "risk avoidance" axis is set according to the level of demand to avoid adverse effects caused by misjudgment of ML components. The method of setting the level of external quality for "risk avoidance" is described in detail below.

1) Procedure for level setting

The flow of setting the level of risk avoidance (AISL) is shown in Figure 2-6.

[41] Setting the level of external quality is a step that corresponds to IEC 61511-1 (JIS C 0511-1), "Process hazard and risk assessment (H&RA)" and "Assignment of safety functions to protective layers.” However, applying the Guidelines does not imply conformance to IEC 61511-1 (JIS C 0511-1).

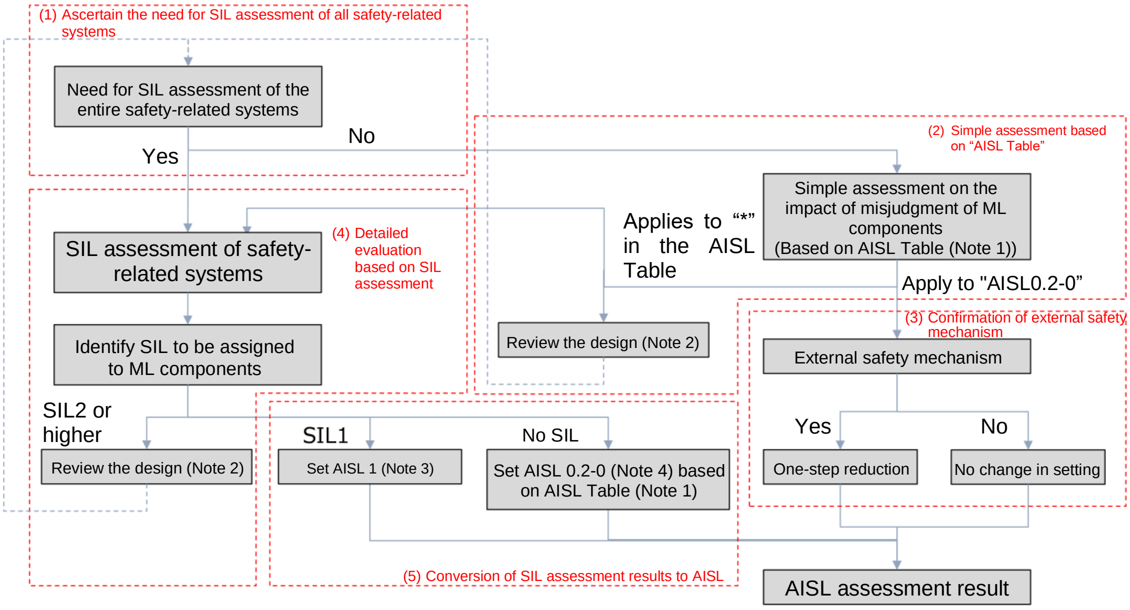


Figure 2-6　Flow of setting the level of risk avoidance (AISL)

Note 1: AISL Table shall mean Table 2-1.

Note 2: When reviewing the design, the assessment should be redone from (1) "Ascertain the need for SIL assessment of all safety-related systems" with a new system configuration and operation (human involvement, etc.) In the Guidelines, the assignment of SIL to ML components shall be set to SIL1 or No SIL by the "Safety-related systems independent of the ML based system" and "external safety mechanism."

Note 3: Conformance to IEC 61508 (JIS C 0508) is required to ensure SIL1 or higher based on SIL assessment. This does not imply that "ML components satisfying AISL1 can be used as SIL1."

Note 4: The AISL corresponding to "no SIL" is "0.2-0." If the SIL to be assigned to ML components is "No SIL" and the AISL assessment based on the AISL Table is "*," priority is given to the "SIL assessment," which is more rigorous. In other words, even if the AISL Table indicates "*," priority is given to "No SIL" based on the SIL assessment, and AISL is set to "0.2." If the AISL Table is assessed to be 0.2-0, the result of AISL Table shall be used.

(1) Ascertain the need for SIL assessment of all safety-related systems

First, the necessity for SIL assessment of the entire safety-related systems to be implemented in the ML based system is confirmed. SIL assessment is a method of defining the safety integrity level (SIL) requirements of safety-related systems as defined in the functional safety standards[42]. By implementing safety-related systems designed based on SIL assessment, it is possible to demonstrate that the safety assurance measures are appropriate. If the equipment on which the ML based system is implemented is assumed to be subject to the functional safety standards, it is judged that SIL assessment is required, and thereby Procedure (4) is taken. In other cases, it is judged that SIL assessment is not required in Procedure (1), and Procedure (2) may be performed.

Even if equipment is subject to the functional safety standards, in the following cases, it is judged that new SIL assessment is not required in Procedure (1), and Procedure (2) is performed.

- When safety of the plant has been confirmed[43] using the methodology of risk assessment in

[42] IEC61508 (JIS C 0508), IEC61511 (JIC C 0511), etc.

[43] Example: "In the case where no dangerous scenarios were found at all in HAZOP," "In the case where machine learning system is not the source of dangerous scenarios and is not involved in propagating the danger in HAZOP,"

accordance with existing safety standards[44]

- When it is considered that the functions of existing safety-related systems and equipment under control (EUC)[45] systems will not be affected by the misjudgment of ML components, because the plant safety is ensured by other safety-related systems (ML based system and independent systems), where reliability has been verified by the methodology of system development process according to functional safety standards[46]

(2) Simple assessment based on “AISL Table”

If it is judged that no SIL assessment is required, a simple assessment of the level of demand for risk avoidance in ML components is performed based on Table 2-1 (AISL Table) (how to use the Table will be described later). If "*" in Table 2 -1 does not apply, perform Procedure (3). If "*" in Table 2-1 applies, demand for risk avoidance in ML components may have become excessive (the level of demand for risk avoidance is too strong to be achieved with the current machine learning and reliability management technologies), and assessment using a simple method may not be appropriate. For this reason, it is necessary that another SIL assessment be required, and Procedure (4) be performed. Or, after reviewing the design and operation of the ML based system, such as securing Safety-related systems independent of the ML based system and increasing human involvement, reassess from (1) “Ascertain the need for SIL assessment of all safety-related systems" to ensure that the risk avoidance requirement does not fall under "*."

(3) Confirmation of external safety mechanism

In the case where AISL 0.2-0 applies and the external safety mechanism[47] does not exist, the level shall be used as the assessment level of AISL as is.

If an external safety mechanism exists, the AISL assessment level[48] shall be the level reduced by one

---

etc. In the case where a safe automatic shutdown of the plant has been established with sufficient reliability, etc.

[44] In addition to the functional safety standards, “IEC 61882: 2016 Hazard and operability studies (HAZOP studies) - Application guide,” “IEC 60812: 2018 Failure modes and Failure modes and effects analysis (FMEA and FMECA),” “JIS C 5750-4-3: 2011 Dependability management - Part 4-3: Analysis techniques for system reliability - Procedures for failure modes and effects analysis (FMEA),” “IEC 61025: 2006 Fault tree analysis (FTA),” “JIS C 5750-4-4: 2011 Dependability management - Part 4-4: Analysis techniques for system reliability - Fault tree analysis (FTA),” etc.

[45] Under the Guidelines, EUC refers to equipment, machinery, devices, and plants used mainly for manufacturing and maintenance.

[46] If it is thought in Procedure (1) that the misjudgment of ML components will not affect safety-related systems and EUC control functions and no new SIL assessment is required, the impact should be reconfirmed in Procedure (2) using Table 2-1. If an impact corresponding to "*" in Table 2-1 is assumed, it is judged that a new SIL assessment is necessary, and Procedure (4) is performed.

[47] It is software or hardware that is processed in parallel or in series with ML components for the purpose of improving safety, that monitors and corrects undesired output of ML components (limits or overwrites the output), and that can be assessed to be sufficiently safe using existing system development process methods that follow functional safety standards such as IEC 61508 (JIS C 0508) and IEC 61511 (JIC C 0511), particularly for software, IEC 61508-3 (JIS C 0508-3), etc. Interlock, etc. are “Safety-related systems independent of the ML based system” and are different from external safety mechanisms. See “1.5 Terminologies” and “Figure 1-2.”

[48] The AISL assessment value shall be 1/0.2/0.1/0. AISL1 corresponds to SIL1, and 0.2-0 correspond to no SIL. The SIL assessment is classified into four levels (4/3/2/1) based on the functional safety standards (IEC61508/JIS C 0508), and the demands for safety functions are specified according to the classification. In the case of "no SIL," there are no special requirements based on the functional safety standards, but standard quality controls are required. However, since machine learning does not have an established method of quality controls like conventional systems, certain guidelines are necessary. For this reason, the level set as "no SIL" is further divided, and the notation of 0.2, 0.1, and 0 is adopted using decimals to maintain the relationship between large and small to seek a certain level of risk avoidance. (Definition is inherited from the “1st edition of Guidelines on Quality Control for Machine Learning”) In the case of "AISL0," there are no special requirements based on the Guidelines; however, standard quality controls are required. (This is not the same as "Not setting the quality of risk avoidance, i.e. no quality control.”)

(e.g. 0.2 to 0.1). However, the external safety mechanism shall monitor and correct (limit or override) undesirable outputs and decisions of the ML components, and shall be designed based on the SIL assessment and operated at all times. [49]The Guidelines state that an external safety mechanism may reduce AISL by appropriately monitoring and correcting (limiting or overriding) the outputs and decisions of ML components if an external safety mechanism exists.

(4) Detailed evaluation based on SIL assessment

If a SIL assessment is judged to be necessary, the SIL assessment of safety-related systems shall be performed according to functional safety standards, and the SIL to be assigned to ML components shall be identified. Identify the SIL at SIL1 or No SIL and proceed to Procedure (5). In the Guidelines, based on the trend of international discussions on functional safety as of March 2021, the assignment of safety functions with SIL2 or higher to ML components shall be prohibited due to the high risk to safety[50]. If the SIL assigned to the ML component is 2 or higher, review the design and reassess from (1) "Ascertain the need for SIL assessment of all safety-related systems" so that the SIL of an ML component is SIL1 or No SIL. When setting numerical targets for the external quality of ML components (e.g. incidence rate of a potentially dangerous misjudgment), the target functional failure scale corresponding to SIL assigned to ML components may be considered as a reference (e.g. in the low-frequency activation request mode of SIL1, the average probability of function failure per activation request is between 10^{-2} and 10^{-1})[51].

(5) Conversion of SIL assessment results to AISL

If the SIL assigned to ML components is 1, AISL is set to 1.

If there is no SIL, AISL is set based on the Table 2-1 (AISL Table). However, provided, that if the assessment result is "*," the result of "SIL assessment," which is more rigorous, shall be applied, and AISL 0.2 shall be set as the most safety-conscious of the AISL's of 0.2, 0.1, and 0 corresponding to "no SIL."

2) "AISL Table"

Table 2-1 (AISL Table) shows the severity of human or economic impact caused by misjudgments of ML components on the vertical axis, and the degree of possibility for humans to avoid misjudgments of ML components on the horizontal axis. This is based on the AISL assessment table described in the "Machine Learning Quality Management Guideline 1st edition with the vertical and horizontal axes modified to reflect the actual situation in the field of plant safety.

---

[49] If the external safety mechanism is software, special attention should be paid to its independence from ML components.

[50] When making future revisions to the Guidelines, they will be updated based on the latest trends in discussions related to functional safety.

[51] The mean probability of functional failure corresponding to SIL specified in IEC 61508-1 (JIS C 0508-1) is used as a reference (e.g. between 10^{-2} and 10^{-1} per activation request in the low-frequency activation request mode of SIL1, etc.) See IEC 61508-1 (JIS C 0508-1) for details. However, when applied to software, IEC 61508-3 (JIS C 0508-3) requires other techniques for each SIL instead of the mean probability of functional failure. This is because it is considered difficult to set the demand for the mean probability of functional failure for software. For this reason, it is assumed that the numerical targets set here will be adjusted through PoC, etc., even at the stage of creating internal quality.

Table 2-1 Criteria for simple assessment of "risk avoidance" (AISL Table)

Severity Criteria (Note 1)	Human Damage	Economic Damage (Direct damage amount only)	Economic Damage (Including indirect damage amount)	Classification of avoidability by humans[52]		
				(1) There is no human alternative system where the results of ML components are directly reflected in operation or maintenance	(2) The judgment results made by ML components are not directly reflected in operation or maintenance, but reflected through human confirmation and application of alternative systems	(3) ML components provide only supplementary information, which is reflected in operation or maintenance through human judgment
I	・ Death ・ Disabling injuries ・ Many serious injuries ・ Extremely large number of casualties	Direct damage amount (Note 4) ≥ 100 million yen	・ Significant impact on the survival, etc. of the corporate entity ・ Serious damage that causes detriment to business operations	*(Note 2)	*(Note 2)	*(Note 2)
II	・ Serious/minor injuries ・ Large number of casualties	Direct damage amount ≥ 10 million yen	Specific damages that cannot be ignored	*(Note 2)	*(Note 2)	AISL 0.2 (Note 5)
III	Micro injury (Note 3)	Direct damage amount < 10 million yen	Only minor loss of profit	*(Note 2)	AISL 0.2 (Note 5)	AISL 0.1
III'	(When it can be easily avoided by assumed victims)			AISL 0.2 (Note 5)	AISL 0.2 (Note 5)	AISL 0.1
IV	No injury assumed	Direct damage amount is insignificant	Economic damage amount, including indirect damage, is assumed to be insignificant	AISL 0	AISL 0	AISL 0

[52] Classification of avoidability in the "Machine Learning Quality Management Guideline 1st edition" interpreted in the field of plant safety.

Note 1: For the severity criteria, select the highest value among "human damage," "economic damage (direct damage amount only)," and "economic damage (including indirect damage amount).” The application of "economic damage (including indirect damage amount)" is optional.

Note 2: In the Guidelines, SIL assessment of the entire safety-related systems is mandatory if “*” applies, and the SIL assigned to ML components should be designed to be SIL1 or No SIL. If "SIL1" has been identified through SIL assessment, "AISL1" shall be set.

Note 3: "Minor injuries" are defined as those with a severity of a so-called "non-serious injuries" or less, and accidents without lost workdays that require medical attention fall under Severity Standard II.

Note 4: "Direct damage amount" shall mean the following. "Cost of repair, replacement, cleaning, disposal, environmental remediation, and emergency response. Direct costs do not include indirect costs such as lost business opportunities, business interruption and lost profits due to loss of raw materials and products, equipment shutdown, temporary equipment procurement and operation costs, and procurement costs of alternative products in response to customer requests. ”[53]

Note 5: If none of the items marked with "*" applies and there is a constantly operating external safety mechanism designed and implemented in accordance with the SIL assessment, the AISL can be reduced by one stage, e.g. 0.2 to 0.1.

Note 6: This Table does not rank the probability of incidence considered in setting SIL based on the risk graph method, etc. in order to make a simple assessment focused on safety. Rather, it treats them as having a uniformly high probability of incidence (i.e. the AISL corresponding to the SIL value is assigned to the one with the highest probability of incidence considered in setting SIL).

[53] Source: Center for Chemical Process Safety (CCPS), "CCPS Process Safety Leading and Lagging Measurement Standards," revised January 2011, translated by SCE-Net Safety Research Group. In the Guidelines, "direct cost" is replaced with "direct damage" and "indirect cost" is replaced with "indirect damage," and they are reflected in the severity criteria.

a. Vertical axis of "AISL Table"

The vertical axis of "AISL Table" indicates that AISL varies according to the severity of human or economic impact that would occur if a midjudgment is made from the ML components. Therefore, the AISL required is set higher at the top of the table where the severity is greater, and lower at the bottom. AISL[54] is set according to this axis. For the severity criteria, the largest damage is selected among "human damage," "economic damage (direct damage amount only)," and "economic damage (including indirect damage amount)." The application of "economic damage (including indirect damage amount)" is optional[55].

If the safety of the plant has been confirmed using a risk assessment method in accordance with existing safety standards[56], the severity shall be set taking into account the risk assessment results using the said method[57]. In addition, if safety-related systems independent of the ML component are used to reduce the impact of a misjudgment by an ML component[58], the severity should be set in consideration of such measures[59].

Table 2-2 summarizes the relationship between the human or economic impacts defined in the Guidelines and the existing assessment criteria and accident classifications in the field of plant safety and the "Machine Learning Quality Management Guideline 1st edition". Please refer to Table 2-1 (AISL Table) when considering the severity standard for the vertical axis.

---

[54] Since the assessment of "each implementation site" is required in this guideline, the reliability assessment is carried out separately for each plant in the case of multiple plants implementation. If the impact of misjudgment could spread to multiple plants, it shall be treated as "the incidence of impacts at a single implementation site is more frequent." In other words, when the effect of misjudgment is examined for each plant, the effect of the same misjudgment is assumed for all plants.[55] The application of "Economic Damage (including indirect damage)" is optional, but if the purpose of introducing an ML based system includes the prevention and reduction of indirect damage such as lost profits due to production stoppage, etc., then it is inevitable that "Economic Damage (including indirect damage)" should be applied to examine the intensity of impact.

[55] The application of "Economic Damage (including indirect damage)" is optional, but if the purpose of introducing an ML based system includes the prevention and reduction of indirect damage such as lost profits due to production stoppage, etc., then it is inevitable that "Economic Damage (including indirect damage)" should be applied to examine the intensity of impact.

[56] In addition to the functional safety standards, "IEC 61882: 2016 Hazard and operability studies (HAZOP studies) - Application guide," "IEC 60812: 2018 Failure modes and Failure modes and effects analysis (FMEA and FMECA)," "JIS C 5750-4-3: 2011 Dependability management - Part 4-3: Analysis techniques for system reliability - Procedures for failure modes and effects analysis (FMEA)," "IEC 61025: 2006 Fault tree analysis (FTA)," "JIS C 5750-4-4: 2011 Dependability management - Part 4-4: Analysis techniques for system reliability - Fault tree analysis (FTA)," etc.

[57] If the risk assessment results cannot be adequately justified, they cannot be considered in the setting of severity.

[58] The reliability of safety-related systems to reduce the impact shall be confirmed by the system development process method in accordance with the functional safety standards. If reliability cannot be adequately justified, the measure cannot be considered in the setting of severity.

[59] If human and direct economic damages are not assumed by Safety-related systems independent of the ML based system, the risk of human and direct economic damages shall be AISL0. Here, if the risk of indirect economic damage remains, it is possible to set a higher AISL and perform quality control based on the risk avoidance axis by assuming the impact including indirect economic damage.

Table 2-2　Relationship between severity standards in the Guidelines and existing standards and classifications

<table>
<tr><th rowspan="3">Criteria of the Guidelines</th><th colspan="4">Existing Assessment Criteria for "Probable Impact"</th><th colspan="3">Existing Assessment Criteria for "Accidents Occurred"</th></tr>
<tr><th colspan="2">”Machine Learning Quality Management Guideline 1st edition”</th><th rowspan="2">High-pressure gas Risk Assessment Guidelines (Ver. 2): Implementation cases of risk assessment</th><th rowspan="2">Occupational safety Guidelines for investigation of danger or hazard, etc.: Attachment 4 Severity of injury/disease</th><th rowspan="2">High-pressure gas Guidelines for responding to accidents at high-pressure gas and oil industrial complexes: Classification of accidents</th><th rowspan="2">Firefighting Severity indicators for fire and spill accidents at dangerous facilities: Indicators for human suffering</th><th rowspan="2">Japan Petroleum Industry Association Accident assessment criteria
(CCPS method)</th></tr>
<tr><th>Human Risk</th><th>Economic Risk</th></tr>
<tr><td rowspan="4">I
Human Damage:
・ Death
・ Disabling injuries
・ Many serious injuries
・ Extremely large number of injuries (serious/minor)
Economic damage (direct damage amount only):
・ Direct damage of 100 million yen or more
Economic damage (including indirect damage amount):
・ Significant impact on the survival, etc. of the corporate entity
・ Serious damage that causes detriment to business operations</td><td>Simultaneous deaths of multiple people</td><td rowspan="2">Significant impact on the survival, etc. of the corporate entity</td><td rowspan="3">I: Death</td><td rowspan="4">(1) Fatal: Accident resulting in death or permanent damage to a body part</td><td rowspan="4">Class A accident
・ 5 or more fatalities
・ Total of 10 or more dead or seriously injured
・ Total of 30 or more dead or seriously injured
・ Direct damage of 500 million yen or more
Class B1 accident (i)
・ 1 to 4 fatalities</td><td rowspan="3">Level 1: Caused death</td><td>Level 1
・ Multiple fatalities
・ Direct damage exceeding 1 billion yen</td></tr>
<tr><td rowspan="2">Single casualty</td><td rowspan="2">Level 2
1 death
・ Direct damage from 100 million to 1 billion yen</td></tr>
<tr><td rowspan="2">Serious damage that causes detriment to business operations</td></tr>
<tr><td>Disabling injuries</td><td rowspan="2">II: Accident with lost workdays</td><td rowspan="2">Level 2: Caused serious/moderate injuries</td><td rowspan="2">Level 3: Accident with lost workdays
・ Direct damage from 10 million to 100 million yen</td></tr>
<tr><td>II
Human Damage:
・ Serious/minor injuries
・ Large number of injuries (serious/minor)
Economic damage</td><td>Severe</td><td>Specific damages that cannot be ignored</td><td>(2) Serious: Accident with lost workdays (lasting more than one month, involving a large number of victims simultaneously)</td><td>Class B1 accident (excluding (i))
・ 2 to 9 persons seriously injured
・ 6 to 29 persons injured
・ Direct damage from 100</td></tr>
</table>

(direct damage amount only): ・ Direct damage of 10 million or more Economic damage (including indirect damage amount): ・ Specific damages that cannot be ignored			III. Accidents without lost workdays	(3) Moderate: (Lasting less than one month, involving multiple victims simultaneously)	million to 500 million yen Class C1 accident (i) ・ 1 to 5 injured and 1 or less seriously injured	Level 3: Caused minor injuries	
III Human Damage: ・ Minor injury Economic damage (direct damage amount only): ・ Direct damage less than 10 million yen Economic damage (including indirect damage amount): ・ Only minor loss of profit	Minor injury	Only minor loss of profit	IV: Minor accident	(4) Mild: Accidents without lost workdays, involving only scratch-level injuries	Class C1 accident (excluding (i))		Level 4 ・ First aid ・ Direct damage from 2.5 million to 10 million yen
III' (When it can be easily avoided by assumed victims)	Minor injury (When it can be easily avoided by assumed victims)					Level 4: No mild injuries	
IV Human Damage: ・ No injury assumed Economic damage (direct damage amount only): ・ Direct damage amount is insignificant Economic damage (including indirect damage amount): ・ Economic damage amount, including indirect damage, is assumed to be insignificant	No injury assumed	No damage assumed	V: No injury		Class C2 accident		Level 5 ・ Below Level 4

b. Horizontal axis of "AISL Table"

The horizontal axis of "AISL Table" indicates that AISL varies according to the avoidability by humans of misjudgments by ML components. (1) is a case where the decision of ML components is final for the entire ML based system as is. (2) is a case where ML components make a decision, but it is always verified by a human to make the final decision. (3) is a case where ML components do not make any decisions themselves (they only output information that can be used to make decisions), and the output of ML components is always reviewed by a human to make the decision. Therefore, the AISL required is set higher at the left of the table, and lower at the right. AISL is set according to this axis.

In considering the horizontal axis, the degree of human involvement in the output and decisions of ML components is confirmed based on Figure 2-7.

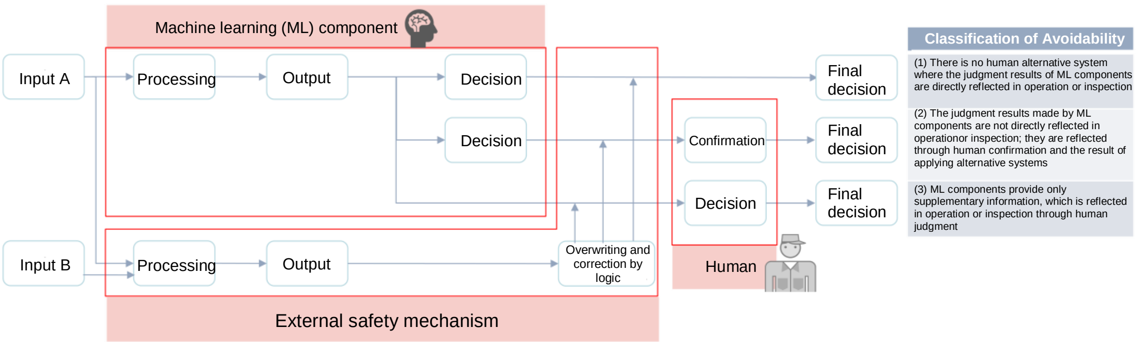


Figure 2-7 Human involvement in output and decision of ML components

(2) Performance

Table 2-3 shows the level setting and assessment criteria of external quality related to "performance." Using the contents of ”Machine Learning Quality Management Guideline 1st edition” formulated by the National Institute of Advanced Industrial Science and Technology (AIST), assessment is based on criteria named AIPL. AIPL is determined by the importance and necessity of adhering to the requirements of a given external quality..

AIPL 2, the highest level, corresponds to the case where it is mandatory or a strong prerequisite for the operation of an ML based system that ML components satisfy certain performance indicators. AIPL 1 corresponds to the case where certain performance requirements are specified as the purpose of the ML based system, but not as strictly as AIPL 2 (i.e. achieving the performance requirements is considered a desirability and not a necessity). AIPL 0 corresponds to the case where no performance indicator is specified, and the purpose of development is to discover the performance indicator itself. The assessment criteria for AIPL 2/1/0 are the same as those given in the ”Machine Learning Quality Management Guideline 1st edition.

Table 2-3 Level setting and assessment criteria for "performance" (AIPL)

Performance Level	Description
AIPL 2 (mandatory requirements)	• When it is mandatory or a strong prerequisite for the operation of an ML based system that ML components satisfy certain performance indicators (e.g. accuracy, precision or recall). • When fulfillment of the abovementioned performance indicators is clearly stated as a requirement in contracts, etc.
AIPL 1 (best-effort requirements)	• When certain performance requirements are specified as the purpose of the ML based system, but they do not fall under AIPL 2. Especially, when adhering to launch schedule is a priority, or when it is permissible to improve the performance gradually through test operation while monitoring the quality.
AIPL 0	• When no performance indicator is specified at the time of development and the purpose of development is to discover the performance indicator itself. • In the case of development that ends at the so-called PoC stage.

The assessment criteria for "performance" in the ”Machine Learning Quality Management Guideline 1st edition” set levels according to "the strictness of requirements for satisfying a certain level of performance (e.g. accuracy, precision or recall)," and do not indicate the level of "the performance itself.” "A higher level is supposed to be required when the level of performance itself is high," but in the criteria of the ”Machine Learning Quality Management Guideline 1st edition”, the level is assessed at the same value regardless of the levels required of performance[60].
The current guidelines set as follows: "AIPL 2" should be applied in the case where "best-effort operation is acceptable, but demanded performance level is high," in order to provide practical "performance" assessment criteria for application to the field of plant safety. [61] The criteria of the levels of performance itself is not set indiscriminately in the Guidelines, and is to be set by mutual agreement between the user and vendor companies.
The concept of AIPL described above is shown in Figure 2-8.

[60] The "Machine Learning Quality Management Guideline 1st edition" does not include the performance standards per se in AIPL indicators, because the specific target values to be achieved vary by application. Since the Guidelines are dedicated to the field of plant safety and so concrete applications can be assumed, the performance standards themselves are included in the AIPL indicators.

[61] The use case "prediction of pipe wall thickness" in the Guidelines gives an example of the application of AIPL2 when the demand for performance itself is high (see 3.3.1 2) b.)
In the practical application example "5 -1. Operation optimization (ENEOS Corporation and Preferred Networks, Inc.)" of this guideline, AIPL2 is set for the external quality "to present operating parameters that maintain the production and energy efficiency indicators above a certain level" when the required level of performance is high.

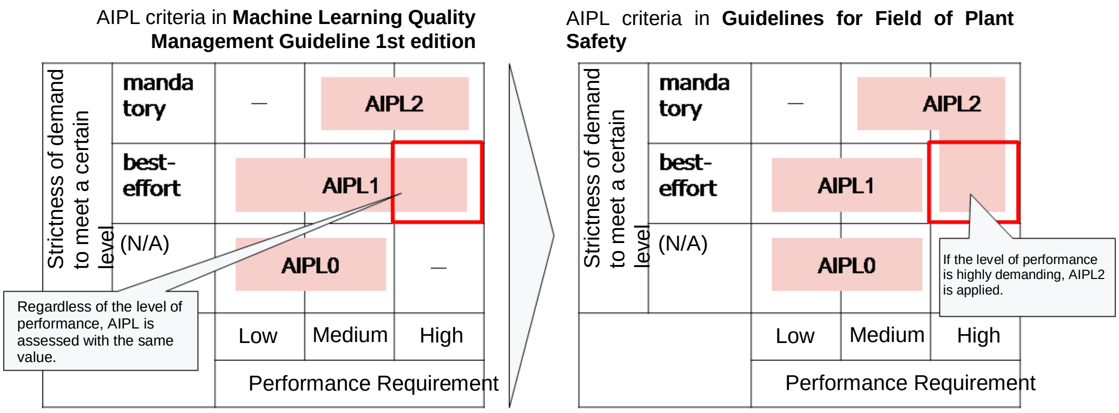


Figure 2-8 Approach to AIPL by "Machine Learning Quality Management Guideline 1st edition" and the Guidelines

2.2.4 Confirming the level of internal quality

External quality whose level was set in the previous sections is achieved by building internal quality. Set the demand levels for internal quality according to AISL/AIPL set in the previous sections[62]. The correspondence between AISL/AIPL of external quality and the demand level of internal quality is shown in Figure 2-9. For example, in the case of AISL 0.2 or AIPL 2, internal quality demands "Level 2." Thus, "Level 2" requirement for each of the eight axes of internal quality is applied. Each level of internal quality is set for each AISL/AIPL, and the highest level of internal quality is used to apply the requirement.

Level of External Quality				
AISL 0.1	**AIPL 1**	O		
AISL 0.2	**AIPL 2**		O	
AISL 1	-			O
Axis of Internal Quality		**Level 1**	**Level 2**	**Level 3**
Sufficiency of requirement analysis		✓ Investigate and record the causes of major quality degradation risks. ✓ ···	✓ Conduct analysis with a certain level of completeness in terms of engineering regarding the deterioration risk of quality in use and its impact on the entire system, and document the results. ✓ ···	✓ Perform the following activities in addition to Level 2. ✓ ···
Coverage for distinguished problem cases		✓ ··· ✓ ···	✓ ··· ✓ ···	✓ ··· ✓ ···
···		✓ ··· ✓ ···	✓ ··· ✓ ···	✓ ··· ✓ ···

Figure 2-9 Correspondence between AISL/AIPL of external quality and demand level of internal quality

[62] Only "Uniformity of datasets" has levels corresponding to AISL and AIPL independently (AISL0.1→LvS1, AISL0.2,1→LvS2, AIPL1→LvE1, AIPL2→LvE2). See "Appendix: 'Uniformity of datasets,' Checklist from the 'Perspectives in the field of plant safety' to ensure internal quality."

2.2.5 Confirming and executing requirements for internal quality

ML components are developed by the requirements based on the level of internal quality identified in the previous section. In doing so, it is necessary to check (1) "Requirements," (2) "Perspectives in the field of plant safety," and (3) "Use case-specific perspectives.” [63]

The eight axes of internal quality and their respective requirements are as stated in the “Machine Learning Quality Management Guideline 1st edition” [(1)].

Furthermore, in order to apply the requirements smoothly to the development of ML components in the field of plant safety, the Guidelines summarize points to be considered specific to the field of plant safety based on actual examples of development in the field [(2)]. In addition, five use cases are set up, to be explained in the following chapter, and the points specific to each use case are also summarized [(3)].

(1), (2), and (3) are summarized in "Appendix: “Perspectives in the Field of Plant Safety” Checklist for Internal Quality Assurance.” When building and operating ML components, refer to the Appendix and meet the requirements designated in "(1) Requirements." Consult "(2) Perspectives in the field of plant safety" and "(3) Use case-specific perspectives." Internal quality requirements are predetermined, but the methods for achieving those requirements (specific efforts to ensure internal quality) vary depending on the external quality (and its corresponding quality in use and functional requirements). In other words, not only does the level of internal quality requirements change according to the level of external quality, but even at the same level, the method of achieving internal quality is different in each individual case.

Internal quality requirements can be considered achieved by addressing all requirements of a given level (Lv 1 – 3), either by meeting the requirements or provide reasons as to why they are not applicable.

Internal quality Lv1 corresponds to AISL0.1 and AIPL1, internal quality Lv2 corresponds to AISL0.2 and AIPL2, and internal quality Lv3 corresponds to AISL1. For example, by addressing all the requirements of internal quality Lv2, it can be explained that AISL0.2 or AIPL2 is ensured.

In this way, hierarchical quality assurance is realized, where “the required level of ‘external quality’ is achieved by improving the ‘internal quality’ of ML components to realize the ‘quality in use’ of the final system"[64].

In the external quality setting stage, a numerical target (e.g. accuracy of more than x%) specific to machine learning is not set. However, regarding the external quality of “risk avoidance,” it may be possible to define numerical targets (e.g. the rate of misjudgments that lead to danger) required for ML components in the process of setting the level of external quality (2.2.3). In addition, in the process of building ML components in this section, both user companies and vendor companies may agree to set specific numerical targets (e.g. accuracy, F-measure) specific to machine learning, depending on the results of PoC, data acquisition, and training status.

The above series of procedures described in Section 2.2 can demonstrate the reliability of an ML component (i.e. that the quality of an ML component is as expected).

In order to demonstrate the reliability of an ML component, the following three points must be satisfied.

---

[63] (1), (2) and (3) are summarized in "Appendix: Checklist from the 'Perspectives in the field of plant safety' to ensure internal quality" of the Guidelines.

[64] National Institute of Advanced Industrial Science and Technology (AIST), “1st edition of Guidelines on Quality Control for Machine Learning”

1) The performance required of an ML component is appropriately limited by non-ML components such as independent safety-related systems and operations such as human involvement
2) Processes for training, testing, implementing, and operating ML components are appropriately designed
3) ML components are sufficiently tested

By setting external quality items and AISL/AIPL according to the procedure, 1) can be accounted for. In addition, 2) can be accounted for by satisfying the internal quality requirements. If the above 1) and 2) are considered in accordance with the procedures described in Section 2.2, and if the required performance is achieved through sufficient testing in 3) above, then all of 1) to 3) would be satisfied.

Column: Does AI need "100% accuracy”?

Machine learning is the process of learning regularities and decision criteria from data, and making predictions and judgment based on them. No matter how much data are collected, it is difficult in principle to achieve 100% accuracy because data are only samples of reality and there are no fixed rules for learned regularities and judgment criteria.

On the other hand, safety is an overarching imperative for plants, and the basic stance for plant owners is to pursue "100% safety" as a principle goal. Therefore, plant owners tend to demand "100% accuracy" in the behavior of ML components as well, which makes it difficult to achieve the development goal and may hinder the project’s progress.

The Guidelines are based on the premise that it is difficult to ensure a high level of safety with ML components alone, and that safety should be ensured by combining ML components with existing systems, so that ML components are not given excessive safety functions.

For example, when introducing an ML based system (use case "equipment deterioration diagnosis (3.3.3)"), which assesses the medium- to long-term degradation trend of equipment in units of several months, the purpose of the system is to optimize the medium- to long-term maintenance, and not “protecting safety by detecting equipment failures with high accuracy.” The safety functions are guaranteed by the existing safety-related systems, and even if the ML based system makes a wrong decision, safety will not be compromised compared to before the introduction of ML components. In this case, as long as the accuracy helps to optimize maintenance, the system can be operated without problems and maintenance can be made more efficient.

In addition, consider the case of introducing an ML based system that presents optimal operating parameters to improve productivity (use case "optimization of operation (3.3.5)"). Here, implementing an “external safety mechanism” which monitors the output of ML compoments and limits the range of operating parameters against non-normal output, or combining ML compoments with preexisting systems e.g. interlocks, will absolve the ML components of the necessity to meet extremely high safety standards.

Furthermore, when introducing an ML based system that detects short-term signs of abnormality in 20–30 minutes or the next few days (use case "detection and diagnosis of early signs of abnormality (3.3.4)"), instead of automatically operating the plant based on an alert of a predicted abnormality, operations can be arranged in a way that instructs a human operator to perform necessary checks before taking actions (including suspension of the plant). This reduces the level of accuracy required of the ML components.

As described above, it is essential to ensure safety not only with ML components, but also by comprehensively including existing safety-related systems, "external safety mechanisms," and operators and maintenance engineers. In this way, the usefulness of machine learning can be appropriately utilized to enhance both safety and efficiency.

In the next chapter, use cases of machine learning in the field of plant safety will be presented with study examples to ensure safety comprehensively, in addition to considering ML components. It is expected that these examples can be used as a reference to devise safety assurance measures for your own cases and to set appropriate accuracy targets for ML components.

3. Use Cases of Machine Learning in the Field of Plant Safety

3.1 Positioning of use cases in the Guidelines

The procedure for reliability assessment presented in the previous section can be applied to overall ML system in the field of plant safety, regardless of how machine learning is used. However, it requires a certain level of familiarity with the Guidelines in order to set specific quality in use and external quality and to implement internal quality requirements smoothly. There is a chance that the process atmay feel complicated, especially when implementing the Guidelines for the first time.
Therefore, "use cases" utilizing the Guidelines are presented for each typical machine learning use case in the field of plant safety, and the items of quality in use/external quality and measures to ensure internal quality are illustrated as examples. This section provides information that can be used as a reference when considering using the Guidelines.

3.2 Scope of use cases

The Guidelines cover three ML based systems for maintenance, "prediction of pipe wall thickness," "pipeline image diagnosis," and "equipment deterioration diagnosis," and two ML based systems for operation, "detection and diagnosis of early signs of abnormality" and "optimization of operation." A summary of each use case is shown in Table 3-1.

Table 3-1 Summary of use case

Use Case	Purpose of Introduction	Function Overview
ML based systems for maintenance		
(1) Prediction of pipe wall thickness	Appropriately timed replacement	Predicting pipe wall thickness from pipe flow rate and contents
(2) Pipeline image diagnosis	Reduction of a visual inspection load	Judge whether or not a visual inspection is needed from the image of piping (screening)
(3) Equipment deterioration diagnosis	Appropriately timed replacement	Predict deterioration of individual equipment parts
ML based systems for operation		
(4) Detection and diagnosis of early signs of abnormality	Avoid operational shutdown due to accidents	Detect signs of abnormality at the plant
(5) Optimization of operation	Improvement of production efficiency and quality	Present the optimal operating parameters for given purposes

Figure 3-1 shows the relationship between the development of an accident, etc. at the plant and the use cases set in the Guidelines. When plant safety is considered as a protective function against the development of an accident, etc. (upper blue lane), the activities in which the protective function is exercised include maintenance activities, operation, response to inappropriate operation, response to abnormality, and emergency response (middle green lane). Although it is considered that there are opportunities to use machine learning in each phase, as there are many cases that aim to use machine learning in maintenance and operation at the time of the introduction of the Guidelines, the five use cases have been set here. [65]

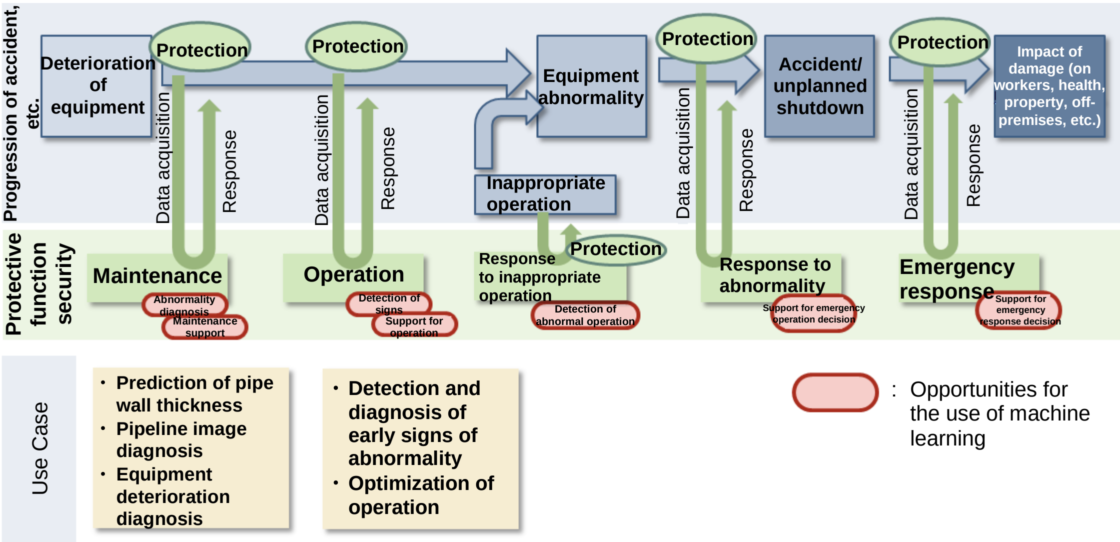


Figure 3-1 Relationship between use cases and progression of an accident, etc. in plants

3.3 Specific application of reliability assessment based on use cases

In this section, the concrete application for each use case is shown in the following structure.
First, as "premise of the use case," the premises of functions and configuration of the ML based system are defined.
Next, as "Examples of quality in use and external quality items," specific items of quality in use and external quality in the use case are presented in risk avoidance and performance axes. In the examples of external quality, if the description assumes the existence of a specific threshold value (e.g. “above a certain level,” “minimize”), determining the value isn't necessary at the stage of setting external quality (2.2.2). Emxaples for such description are as follows: “(Maintain the percentage of correct judgment) above a certain level,” “Minimize (the percentage of false positive).” As for the external quality of "risk avoidance," it may be possible to define numerical targets required for ML components (e.g. incidence rate of a potentially dangerous misjudgment) by checking safety-related systems and external safety mechanisms in the process of setting the level of external quality (2.2.3). In addition, at the stage of building ML components (2.2.5), specific numerical targets (e.g. accuracy, F-measure) unique to machine learning are ultimately set according to the results of PoC and the status of data acquisition and learning, upon mutual agreement between the user and vendor companies.
Finally, as “’use-case-specific perspectives’ to ensure internal quality,” points of consideration are provided specific to each use case regarding development of ML components compliant of internal quality requirements.
The points of consideration that are commonly applicable to the field of plant safety regardless of use case are summarized in "Appendix: “Perspectives in the Field of Plant Safety” Checklist for Internal Quality Assurance.” The Checklist presents an overview of items (requirements and perspectives for each use case) listed in the Guidelines related to meeting internal quality.
At the beginning of each use case, case studies in the field of plant safety are provided to further understanding of each use case.

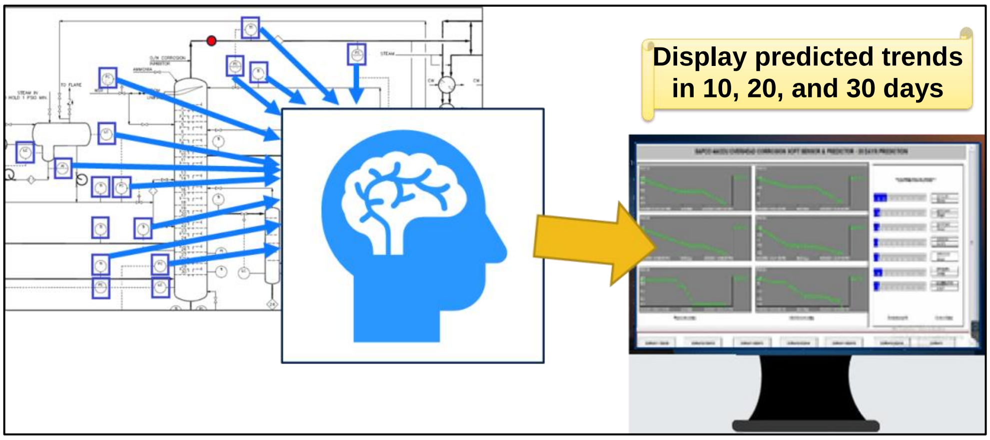
[65] In the future revision of the Guidelines, additional use cases will be considered based on the progress in the use of the ML based systems in the field of plant safety.

In the Guidelines, the setting of AISL and AIPL for external quality and the requirement levels of internal quality are provided only up until the items of external quality; no specific requirement levels are set in the Guidelines. The setting of AISL, AIPL and the requirement levels of internal quality are to be set by the reader, based on the premises and conditions of each ML based system in question.
Use cases in the Guidelines are based on several assumptions, including about ML models, and serve only as examples demonstrating possible viewpoints for achieving quality in use, external quality, and internal quality requirements. Therefore, it is necessary for the readers to compare the use case with the functional requirements and the configuration of their own ML based system, and to apply the examples of use case flexibly.

3.3.1 Prediction of pipe wall thickness

*This case study is provided as a conceptual image of "prediction of pipe wall thickness" and is not a premise for describing subsequent use cases.

Related case study: "Estimation of pipe wall thinning due to corrosion" (Yokogawa Electric Corporation)[66]



The upper piping of an atmospheric distillation column used in oil refineries inevitably degrades from corrosion. By visualizing the progression of corrosion, it is possible to improve the efficiency of maintenance and minimize further development of corrosion by adjusting the operation.

Periodic measurements of wall thickness at 20 locations in a span of two years, as well as process data related to the upper piping of the distillation column, are used as training data. The relationship between process data and wall thinning is modeled using a regression model (via supervised learning). By displaying the estimated amount of wall thinning in real time on the operator's screen, the timing of maintenance can be optimized. Furthermore, it allows operators to be aware of the condition of the piping during operation.

1) Premise of the use case

a. Overview

The use case "prediction of pipe wall thickness" is an ML based system that predicts the current pipe wall thickness based on process data, etc. The purpose of this system is to ensure safety by detecting the progress of wall thinning based on prediction, and at the same time, to improve maintenance efficiency by reducing unnecessary inspection and replacement.
Currently, pipe maintenance due to wall thinning is mainly performed by Time Based Maintenance (TBM). However, there are cases where corrosion progresses rapidly between periodic inspections,

[66] This case study is described in detail in the "Collection of Case Examples of Leading Companies Introducing AI into Plants - Practical Examples of Achieving Results and Breaking Through Challenges in AI Projects -." Consulting the collection is advised. Note that the collection is available in Japanese only.

as well as cases where a large amount of lost profit is incurred due to inspecting and replacing piping that does not yet require maintenance. By shifting to Condition Based Maintenance (CBM) based on the prediction of wall thickness, it is expected that safety will be enhanced, and lost profit reduced.

b. Functional requirements

In this use case, "Predicting pipe wall thickness" is set as a functional requirement of the ML based system.

c. Conceptual image of introduction

In this use case, a conceptual image of introduction is set as shown in Figure 3-2. In conventional maintenance of plant piping, the installed pipes are regularly checked by the maintenance engineer at a predetermined period (TBM)[67]. On the other hand, with the ML based system, the ML components predict the current pipe wall thickness in real time and provide wall thickness predictions to maintenance engineers. The maintenance engineers use this output as well as various sensor data to determine whether or not an actual measurement of wall thickness by maintenance engineers should be performed.

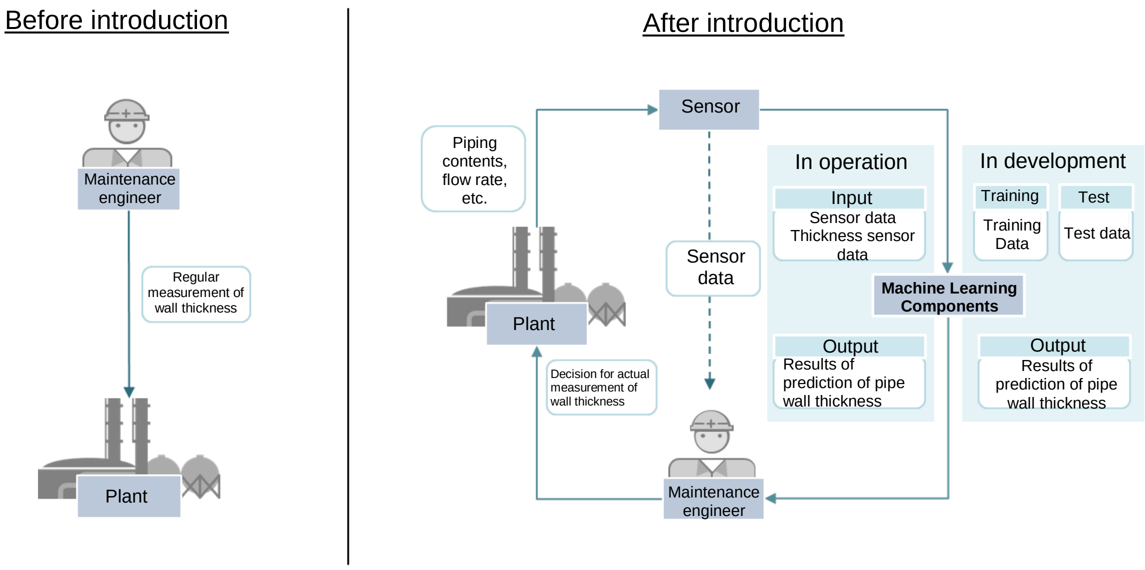


Figure 3-2 Conceptual image example: Introduction of use case “prediction of pipe wall thickness”

d. Relationship with other systems

In this use case, the relationship between the ML components and other systems is set as shown in Figure 3-3. ML components input data such as pipe contents and flow rate, predict the wall thickness of the pipes, output the predicted wall thickness value, and present the predicted wall thickness

[67] Currently, many plant owners have their workers measure the actual wall thickness several times a year (from once a month to once a year) depending on the risk of the piping and other factors. As for legal obligations, for example, the High Pressure Gas Safety Act and its related regulations require wall thicknesses to be measured once a year, and more frequent measurements are regarded as a voluntary safety activity.

value to maintenance engineers. Maintenance engineers determine whether the predicted wall thickness value exceeds the threshold value requiring maintenance, and decide whether maintenance is required, taking into account the data collected by various sensors and other conventional means.

There is no external safety mechanism to monitor and correct the output of ML components (predicted wall thickness value). It is also assumed that there are no Safety-related systems independent of the ML based system to prevent the contents of the piping from leaking.

Therefore, under the premise of this use case, column (3) will be applied when referring to the "AISL Table."

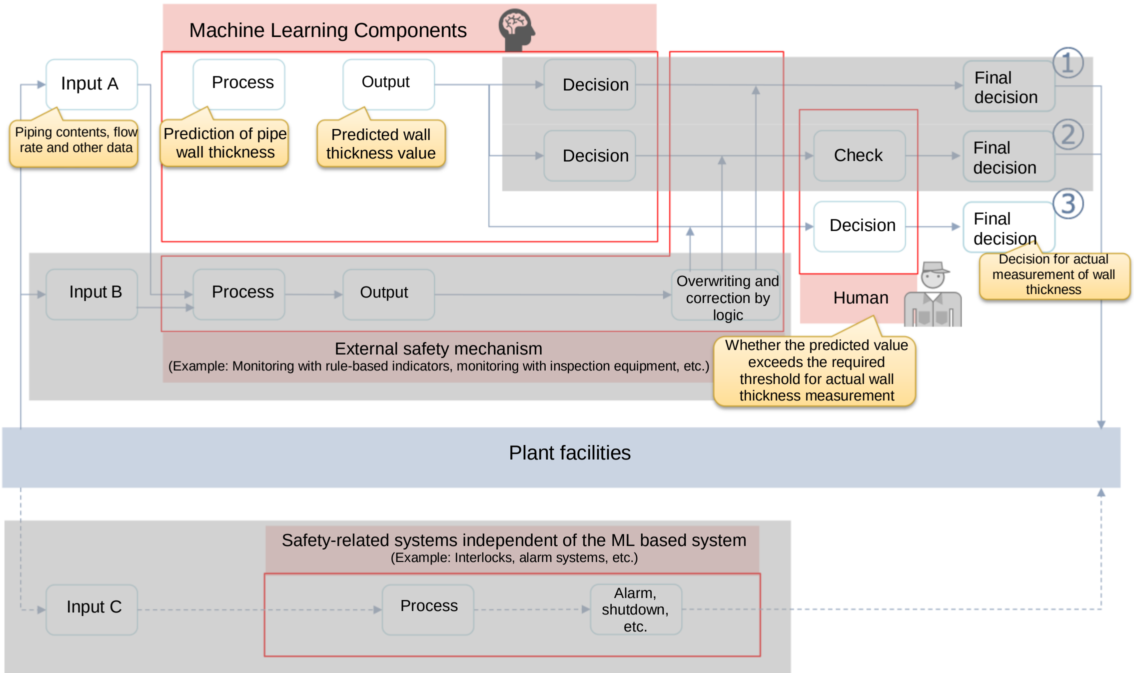


Figure 3-3 Example: Relationship between ML components and other systems in the use case “prediction of pipe wall thickness”

e. Composition of ML components

In this use case, the composition of ML components is set as shown in Table 3-2. In order to predict the value of wall thickness, a supervised regression model is assumed as a model learning the relationship between the wall thickness value and the type of contents, flow rate, flow velocity, pipe material, pressure, etc. that are considered to affect the degree of thinning of the pipe wall. As training data, various pieces of data that affect the degree of pipe wall thinning and the actual measured values of wall thickness are used, and the same data are used for test data.

Table 3-2 Example: Composition of ML components for “prediction of pipe wall thickness”

Learning methodology	Regression (supervised)
Learning model	Learn the relationship between the wall thickness value and the contents, flow rate, flow velocity, material, pressure, etc., of piping
Input data in operation	Piping contents, flow rate, flow velocity, material and pressure data
Training data in	Piping contents, flow rate, flow velocity, material, pressure

development	data, thickness sensor data (actual measured values)
Test data in development	Piping contents, flow rate, material, flow velocity, pressure data, thickness sensor data (actual measured values)

2) Examples of items for quality in use and external quality

Assuming that this use case is set, quality in use and external quality items are set according to Table 3-3. These correspond to "(1) Set what is required to be achieved by the system using machine learning" and "(2) Set the required output for ML components and determine the level of achievement" in the hierarchical quality assurance procedure shown in Chapter 2 (cf. Figure 3-4).

Table 3-3 Example: Quality in use and external quality items for the use case "prediction of pipe wall thickness"

Quality in Use	External Quality
Risk avoidance	
Do not overlook piping that requires actual wall thickness measurement by a maintenance engineer (S-U1)	Keep errors in predicting wall thickness thicker than it actually is within certain limits (S-E1).
Performance	
Eliminate unnecessary maintenance (P-U1)	Keep errors in predicting wall thickness thinner than it actually is within certain limits (P-E1)

Note: Codes for each item are given for convenience of explanation to clarify the relationship between items, and are not required to be given as the Guidelines, nor are they related to the level of the item. (S: Safety, P: Performance, U: Use, and E: External)

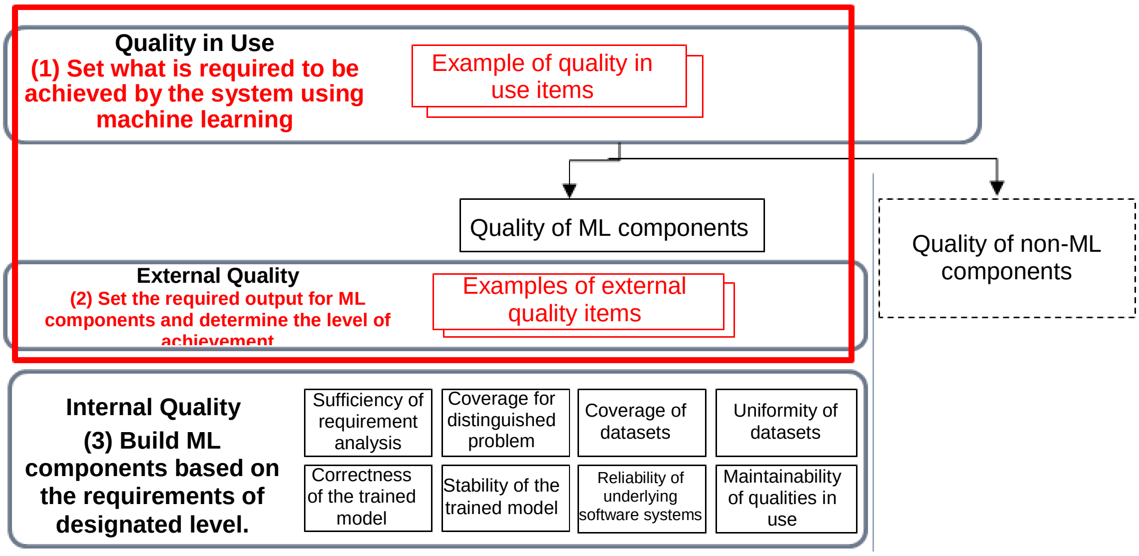


Figure 3-4 Positioning of quality in use/external quality items in hierarchical quality assurance

a. Example of "risk avoidance" Quality

In light of the functional requirement "Predicting pipe wall thickness," the quality in use and external quality shall be set specifically. This is a measure of quality control, from the perspective of “risk avoidance,” in order to prevent human and economic loss due to failing to achieve the functional requirement.

- Quality in Use

  From the perspective of preventing human and economic damage, it is undesirable to make a mistake in predicting the wall thickness that results in the wall thickness actually falling below the threshold value for replacement. In order to minimize this risk, the quality in use of the "risk avoidance" attribute should be set to "Do not overlook piping that requires actual measurement of wall thickness by a maintenance engineer (S-U1).”

- External Quality

  For the set quality in use, determine the external quality required for the output of ML components. For external quality corresponding to the quality in use, "Do not overlook piping that requires actual measurement of wall thickness by a maintenance engineer (S-U1)," “Keep errors in predicting wall thickness thicker than it actually is within certain limits (S-E1)” is set. The phrase "error in predicting wall thickness thicker than it actually is" is used here instead of "error in predicting the actual wall thickness." This is because the risk is not simply that the prediction will be wrong, but also that the prediction will be thicker than the actual value. In this case, while the timing of maintenance is delayed, wall thinning will progress decisively and the contents will leak, causing human and economic losses. On the other hand, "error in predicting wall thickness thinner than it actually is" is also considered, but this is set as a quality in the "performance" axis later.

  For this external quality, the required external quality level "AISL" is set. For “Keep errors in predicting wall thickness thicker than it actually is within certain limits (S-E1),” consider the magnitude of human and economic damage that can be expected if the wall is predicted to be thicker than it actually is, and set the AISL according to the criteria. The set AISL becomes the AISL of ML components, and the required level of internal quality is determined accordingly.

b. Example of "performance" Quality

In light of functional requirement "Predicting pipe wall thickness," in order to control quality to achieve functional requirements at a desirable level from the perspective of "Performance," quality in use and external quality shall be specifically set.

- Quality in Use

  In predicting pipe wall thickness, the frequency of maintenance, such as actual measurement of wall thickness by workers, shall be set to an appropriate level (i.e. not more than necessary). If the frequency of maintenance is higher than necessary, the cost of maintenance and opportunity loss will be significant. Therefore, it is reasonable to minimize the frequency of replacement, assuming that safety is ensured by the quality in use of the "risk avoidance" axis. Therefore, the item "Eliminate unnecessary maintenance (P-U1)" shall be set as quality in use.

- External Quality

  For the set quality in use, determine the external quality required for the output of ML components. For external quality corresponding to the quality in use, "Eliminate unnecessary maintenance (P-U1)," “Keep errors in predicting wall thickness thinner than it actually is within certain limits (P-E1)” is set. What is considered by plant maintenance managers as “performing maintenance work more than necessary” is restated as “predicting wall thickness as thicker than in reality,” using the output of ML component.

  For this external quality, the required external quality level "AIPL" is set. For "Keep errors in predicting wall thickness thinner than it actually is within certain limits (P-E1)," the level of the required accuracy rate and the degree to which it is essential are examined, and the AIPL is set according to the criteria. AIPL2 may be applied to the maintenance and replacement of piping, as necessary, to meet the stringent requirements to save maintenance costas much as possible (assuming that safety is ensured by "risk avoidance").

3) “Use case-specific perspectives” to ensure internal quality

Based on the premises of this use case, points to consider ("perspective") for fulfillingthe requirements of each internal quality are shown in Table 3-4. The “perspectives” described below can be used as a reference when developing ML components similar to this use case. These requirements correspond to “(3) Build ML components based on the requirements of designated level” in the hierarchical quality assurance procedure shown in Chapter 2 (cf. Figure 3-5).

Table 3-4 “Use case-specific perspectives” pertaining to “prediction of pipe wall thickness”

* See the Checklist in Appendix for the list of internal quality requirements and perspectives for this case

Internal Quality	Requirement[68]	Use case-specific perspectives
Sufficiency of requirement analysis	(Common requirements)	・ As a type of corrosion affects the development of "Coverage for distinguished problem cases" and "Coverage of datasets," narrow down the scope to the specific type of corrosion.
Coverage for distinguished problem cases	・ (Lv1) Furthermore, extract attributes of differences in particularly-important environmental factors and prepare cases corresponding to combinations with serious risk factors.	・ "Environmental factors" in this context refer to climate, salinity (regional characteristics such as distance from the sea and wind direction), and others.

[68] In the table, only the requirements related to "Use case-specific perspectives" are excerpted from the "1st edition of Guidelines on Quality Control for Machine Learning." Items that are not listed in this table are also included in the requirements.

Internal Quality	Requirement[68]	Use case-specific perspectives
Coverage of datasets	・ (Lv1) Consider the source and method of acquiring test datasets to ensure that no bias is found in application situations.	・ "Application status" in this context refers to the targeted piping and the frequency of observations, the time axis of evaluation (e.g. whether to make a real-time projection), and others.
	(Common requirements)	・ Pay attention to whether the range of data of assumed attributes, such as contents, flow rate, material, flow velocity and pressure of piping, is covered.
Uniformity of datasets	(Common requirements)	・ Pay attention to ensure that the amount of data for each range of data to be covered by the abovementioned attributes is sufficient. ・ When the amount of data in a specific range is not sufficient, keep in mind that the prediction accuracy of that range could become low.
Correctness of the trained model	-	-
Stability of the trained model	-	-
Reliability of underlying software systems	-	-
Maintainability of qualities in use	(Common requirements)	・ Based on the judgment of whether or not replacement is necessary using the existing method, the actual condition of piping at the time of replacement, and others, verify the actual accuracy and the presence of oversight. ・ As maintaining accuracy is vital in this case, it is crucial to keep extensive records of background information on model construction and types of training data. Consult these records every time changes are to be made after operation begins.

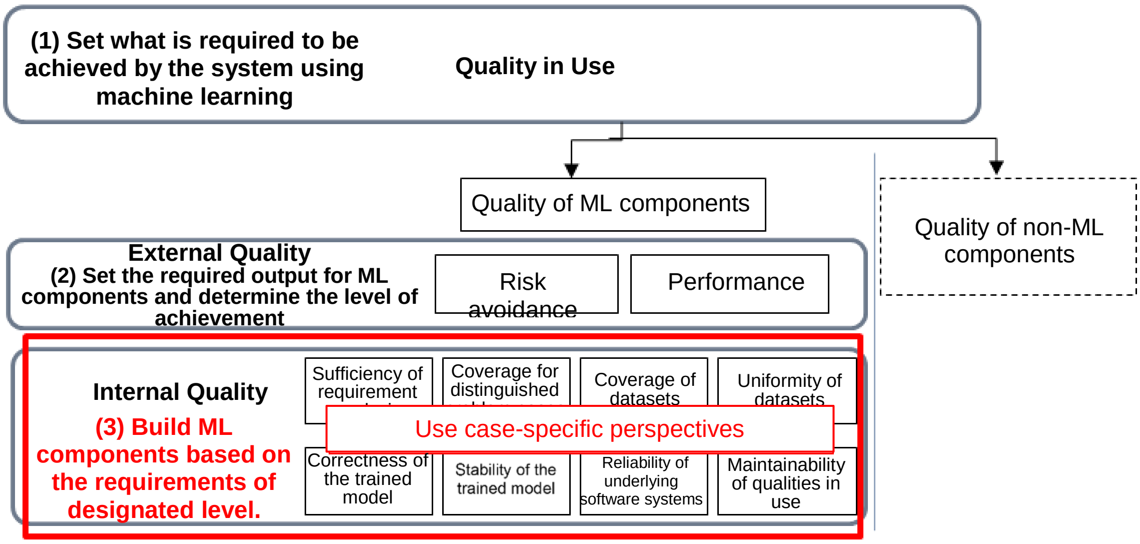
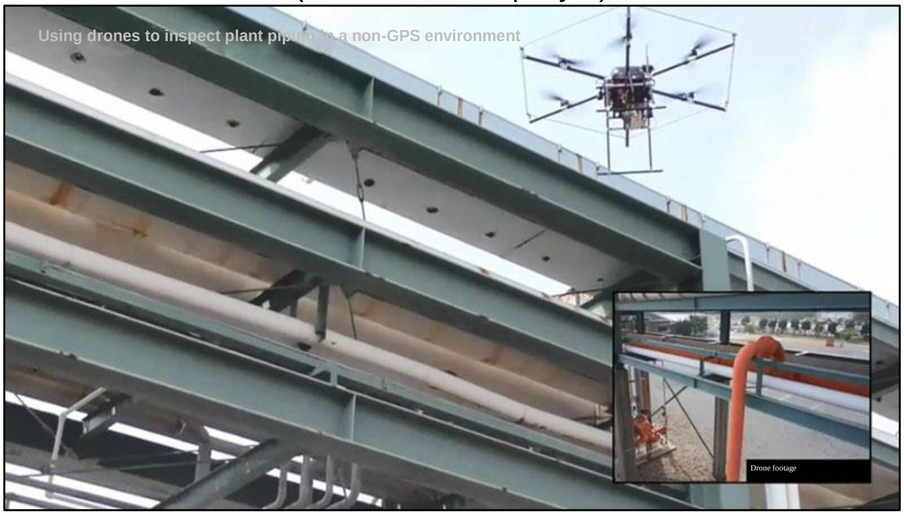


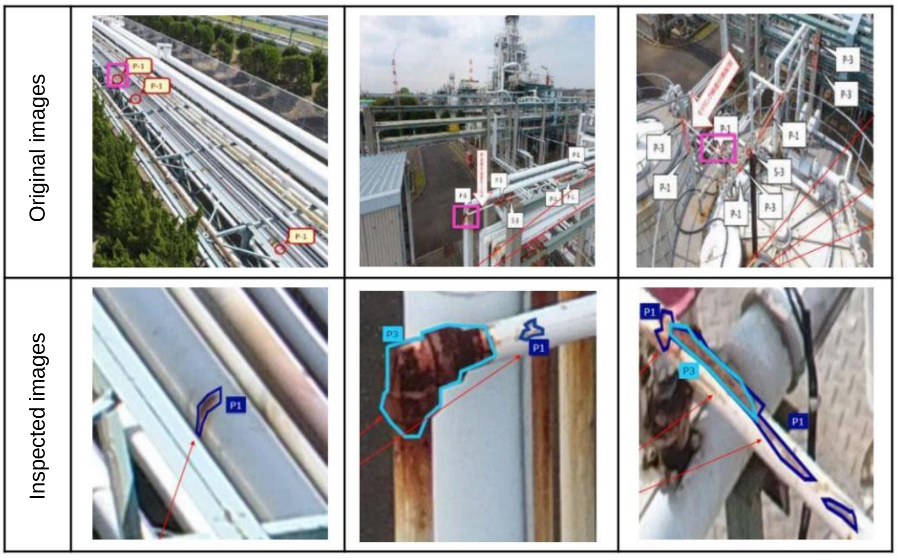
Figure 3-5　Positioning of "Use case-specific perspectives" in hierarchical quality assurance

3.3.2 Pipeline image diagnosis

*This case study presents a conceptual image of "Pipeline image diagnosis" and is not a premise for description in the subsequent use case.

Related case study: “Extraction of corroded areas on the outer surface of pipes by automatic image inspection”
(Plant owner company X)





Corrosion on the outer surface of pipes is detected in periodic inspections by engineers, which is a heavy workload. Especially for piping in high places, scaffolding is required, making frequent inspections difficult. Therefore, this system combines the technology to screen corrosion areas by machine learning and the technology of taking images of pipes with a drone to reduce the inspection workload, expand the inspection range, and increase the frequency of inspections.

The system uses a supervised classification model based on the past image data of external pipe surfaces (labeled with no corrosion or with corrosion). Imagesof external surfaces of pipes are used as input, and the images in which pipes are judged to have corroded are presented to maintenance engineers as screening results. After the images are checked by a maintenance engineer, the corrosion status is confirmed in the field, and a decision is made on whether or not an action is required.

1) Premise of the use case

a. Overview

The use case "pipeline image diagnosis" is an ML based system to screen the areas that need to be visually inspected by a maintenance engineer in order to reduce the burden of inspection work for external pipe surfaces. In the past, the corrosion inspection of the external pipe surface was carried out visually by a maintenance engineer on a regular basis, and this required a great deal of labor. If the screening can limit the number of areas to be visually inspected by a maintenance engineer, it will greatly reduce their workload. In addition, by reducing the workload of experienced maintenance engineers, the facility’s safety can be maintained over thelong term.
This use case is not intended to limit the means of capturing images, but in the case of use in combination with drone photography, as is the case in the case study shown above, it is expected to have an even greater cost-saving effect and improve safety. This is because drones can be used to inspect pipes in high places more frequently, which would typically require scaffoldings.

b. Functional requirements

In this use case, "Determine whether or not a visual inspection is needed based on the images of piping" is set as a functional requirement of the ML based system.

c. Conceptual image of introduction

In this use case, a conceptual image of introduction is set as shown in Figure 3-6. Traditionally, the presence of corrosion in pipes was found out by the visual inspection of maintenance engineers on a regular basis. With the introduction of the ML based system, the ML components identify corrosions that require a visual inspection by a maintenance engineer, and output the judgment to the maintenance engineer. The maintenance engineer checks the image data of relevant areas and decides whether the pipe should be repaired or replaced based on the visual inspection result at the site.

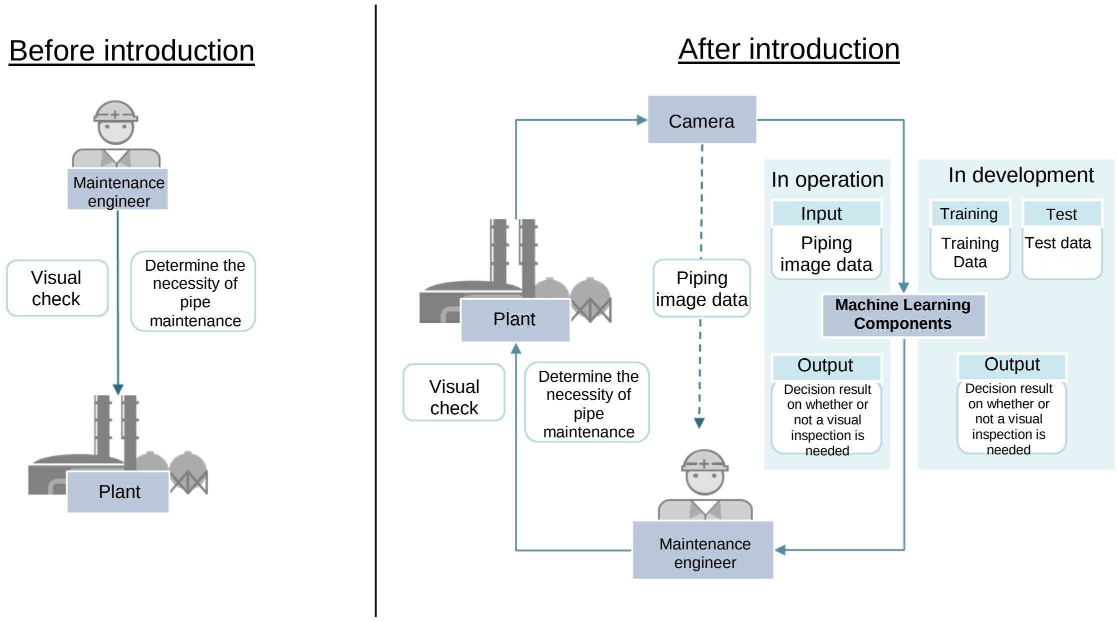


Figure 3-6 Conceptual image (example): Introduction of use case “pipeline image diagnosis”

d. Relationship with other systems

In this use case, the relationship between the ML components and other systems is set as shown in Figure 3-7. ML components use the pipe image data as input to determine whether or not a visual inspection is needed, and output the judgment result to the maintenance engineer. Because the output of ML components in this use case is intended to screen whether or not a visual inspection is needed, the subsequent flow will be branched depending on the judgment. If a visual inspection is deemed not necessary for a given image, the final decision result is that a visual inspection will not be conducted. If a visual inspection is judged to be necessary, a maintenance engineer will check the image to determine whether a visual inspection is actually necessary, and make that judgment as a final decision. It is assumed that there is no external safety mechanism that monitors and corrects the output of ML components per se (whether or not a visual inspection is needed). For this reason, when referring to the "AISL Table," it is necessary to apply column (1) where a higher level is set. It is also assumed that there are no Safety-related systems independent of the ML based system to prevent the contents of the piping from leaking.

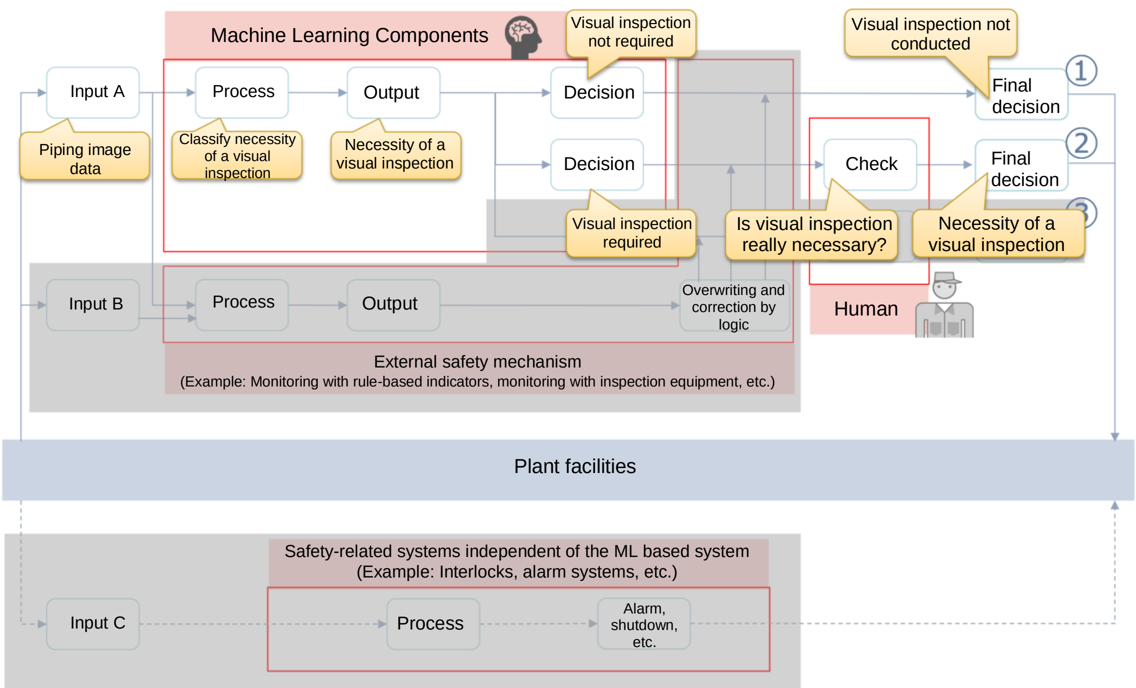


Figure 3-7 Example: Relationship between ML components and other systems in the use case “pipeline image diagnosis”

e. Composition of ML components

In this use case, the composition of ML components is set as shown in Table 3-5. A collection of pipe images from past periodic pipe maintenance activities is utilized, where corrosions in each iamge are given a label indicating whether a visual inspection is necessary. The use case assumes a classification model trained through supervised learning, which would use the aforementioned image data as a basis to output whether a visual inspection is necessary. Pipe image data and labels for the pipe images indicating whether or not a visual inspection is needed are used as training data. Test data are constructed in a structure identical to training data.

Table 3-5 Example: Composition of ML components for “pipeline image diagnosis”

Learning methodology	Classification (supervised)
Learning model	Classification model that classifies whether or not a visual inspection is needed based on the piping image characteristics
Input data in operation	Piping image data
Training data in development	Piping image data + labels indicating whether a visual inspection is necessary
Test data in development	Piping image data + labels indicating whether a visual inspection is necessary

2) Examples of items for quality in use and external quality

Assuming that this use case is set, quality in use and external quality items are set according to Table 3-6. These correspond to “(1) Set what is required to be achieved by the system using machine learning” and “(2) Set the required output for ML components and determine the level of achievement” in the hierarchical quality assurance procedure shown in Chapter 2 (cf. Figure 3-8).

Table 3-6 Example: Quality in use and external quality items for the use case “pipeline image diagnosis”

Quality in Use	External Quality
Risk avoidance	
Do not overlook piping that requires a visual inspection (S-U1)	When a visual inspection is "required," reduce the false-negative rate as much as possible where it is judged to be "not required" (S-E1)
Performance	
Reduce the number of a visual inspections conducted by a maintenance engineer (P-U1)	When a visual inspection is "not required," reduce the rate of false positivewhere it is judged to be "required" within a certain range (P-E1)

Note: Codes for each item are given for convenience of explanation to clarify the relationship between items, and are not required to be given as the Guidelines, nor are they related to the level of the item. (S: Safety, P: Performance, U: Use, and E: External)

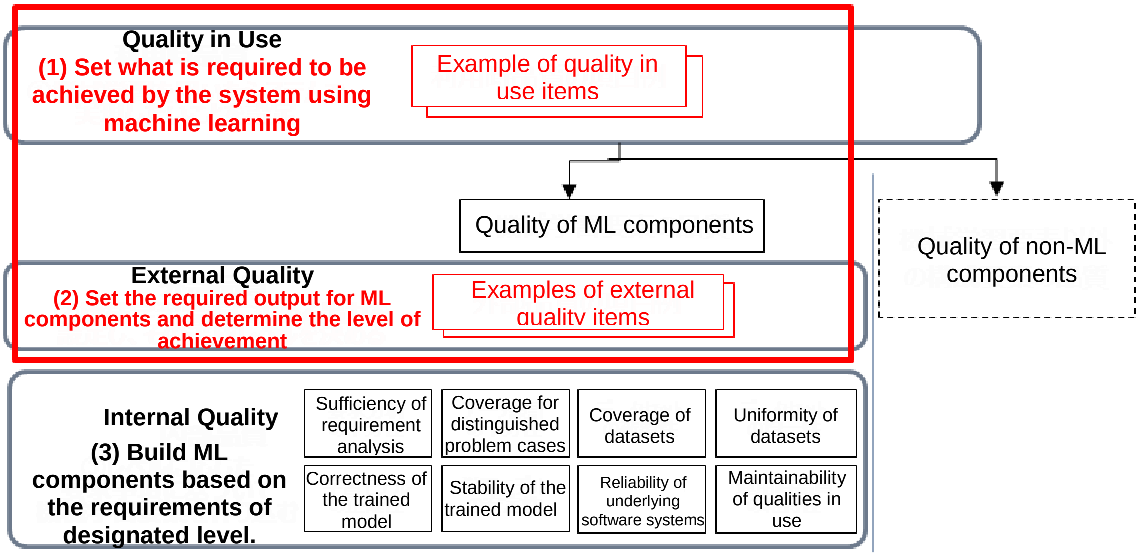


Figure 3-8 Positioning of quality in use/external quality items in hierarchical quality assurance

a. Example of "risk avoidance" Quality

In light of the functional requirement "Determine whether or not a visual inspection is needed based on the images of piping," in order to control quality to prevent human and economic damage due to non-achievement of the functional requirements from the perspective of "risk avoidance," the quality in use and external quality shall be specifically set.

- Quality in Use

  From the perspective of preventing human and economic damage, it is undesirable for pipes requiring a visual inspection to be overlooked. In order to minimize this risk, the quality in use of the "risk avoidance" attribute should be set to "Do not overlook piping that requires a visual inspection (S-U1)."

- External Quality

  For the set quality in use, determine the external quality required for the output of ML components. For external quality corresponding to the quality in use, "Do not overlook piping that requires a visual inspection (S-U1)," "When a visual inspection is 'required,' reduce the false-negative rate as much as possible where it is judged to be 'not required' (S-E1)" is set. For this external quality, the required external quality level "AISL" is set. For "When a visual inspection is 'required,' reduce the false-negative rate as much as possible where it is judged to be 'not required' (S-E1)," consider the magnitude of human and economic damage that could be expected if a piping image that requires a visual inspection is mistakenly judged to be not required, and set the AISL according to the criteria. If ML components judge that a visual inspection is not required, the maintenance engineer does not perform a visual inspection. Therefore, the AISL Table should be set to correspond to the criterion that "There is no human alternative system, and the results of ML components' decisions are directly reflected in operation or inspection." The set AISL becomes the AISL of ML components, and the required level of internal quality is determined accordingly.

b. Example of "performance" Quality

In light of functional requirements "Determine whether or not a visual inspection is needed based on the images of piping," in order to control quality to achieve functional requirements at a desirable level from the perspective of "Performance," quality in use and external quality to be controlled shall be specifically set.

- Quality in Use

  In pipeline image diagnosis, the frequency of a visual inspection must be set at an appropriate level. If the frequency of a visual inspection is higher than necessary, it would be undesirable as the effect of labor reduction cannot be obtained. Therefore, the item "Reduce the number of visual inspections conducted by maintenance engineers (P-U1)"[69] shall be set as quality in use.

- External Quality

  For the set quality in use, determine the external quality required for the output of ML components. For external quality corresponding to the quality in use, "Reduce the number of visual inspections conducted by maintenance engineers (P-U1)," “When a visual inspection is ‘not required,’ reduce the rate of false positive where it is judged to be ‘required’ within a certain range (P-E1)” is set. The concept of “performing more visual inspections than necessary” as recognized by plant maintenance engineers is translated into output of ML components, expressed as “A false positive where a visual inspection is judged as necessary when in reality it is not required”.
  For this external quality, the required external quality level "AIPL" is set. For “When a visual inspection is ‘not required,’ reduce the rate of false positive where it is judged to be ‘required’ within a certain range (P-E1),” the level of required accuracy rate and the degree to which it is essential are examined, and the AIPL is set according to the criteria.

3) “Use case-specific perspectives” to ensure internal quality

Based on the premises of this use case, points to consider ("perspective") for fulfilling the requirements of each internal quality are shown in Table 3-7. The “perspectives” described below can be used as a reference when developing ML components similar to this use case. These requirements correspond to “(3) Build ML components based on the requirements of designated level” in the hierarchical quality assurance procedure shown in Chapter 2 (cf. Figure 3-9).

---

[69] In the practical example of this guideline "2. Pipeline image diagnosis (Mitsubishi Chemical Corporation and NEC Corporation)", quality is set as follows. Please also refer to it.
Quality in use "Do not judge parts that do not require detailed inspection as requiring inspection"
External Quality "When the degree of corrosion is judged with the 3 levels of A/B/C, keep false positives that C is judged to be A/B within certain limits."[70]In the table, only the requirements related to "Use case-specific perspectives" are excerpted from the "1st edition of Guidelines on Quality Control for Machine Learning." Items that are not listed in this table are also included in the requirements.

Table 3-7 "Use case-specific perspectives" pertaining to "pipeline image diagnosis"

* See the Checklist in Appendix for the list of internal quality requirements and perspectives for this case

Internal Quality	Requirement[70]	Use case-specific perspectives
Sufficiency of requirement analysis	(Common requirements)	・ When handling piping that is wrapped with insulation, be aware that the deterioration of insulation will be the target, not the deterioration of the piping itself.
Coverage for distinguished problem cases	・ (Lv1) Furthermore, extract attributes of differences in particularly-important environmental factors and prepare cases corresponding to combinations with serious risk factors.	・ "Environmental factors" in this context refer to sunlight, weather, seasons, time of day, and others.
	(Common requirements)	・ As the color of the piping itself may be covered due to painting or anti-rust painting, ensure the accuracy by considering these differences. ・ Keep in mind that there are cases where it is not possible to directly confirm the outer surface of the piping by images, such as when there is snow accumulation on the piping. ・ Consider keeping the data quality at a certain level by establishing rules and points to consider for photographing. ・ Blurred images may be included in training dataset so that the model can make decisions for such images as well. In such case, potential increase in system complexity and uncertainty should be put into consideration.
Coverage of datasets	(Common requirements)	・ Consider measures to deal with blurred input images, such as image blurring due to the surrounding environment (e.g. sunlight, time) or in drone photography. ・ Pay attention to whether the data range of each attribute of environmental factors is covered.

[70]In the table, only the requirements related to "Use case-specific perspectives" are excerpted from the "1st edition of Guidelines on Quality Control for Machine Learning." Items that are not listed in this table are also included in the requirements.

Internal Quality	Requirement[70]	Use case-specific perspectives
Uniformity of datasets	(Common requirements)	· Pay attention to ensure that the amount of data for each range of data to be covered by the abovementioned attributes is sufficient. · When the amount of data in a specific range is not sufficient, keep in mind that the prediction accuracy of that range could become low.
Correctness of the trained model	-	-
Stability of the trained model	-	-
Reliability of underlying software systems	-	-
Maintainability of qualities in use	(Common requirements)	· Verify accuracy using an image photographed during the operation phase. Record the results of visual inspections performed upon AI's decision, and compare these results to the results of aforementioned verification. If the judgment accuracy is low, perform a thorough check on input image and model used in the verification.

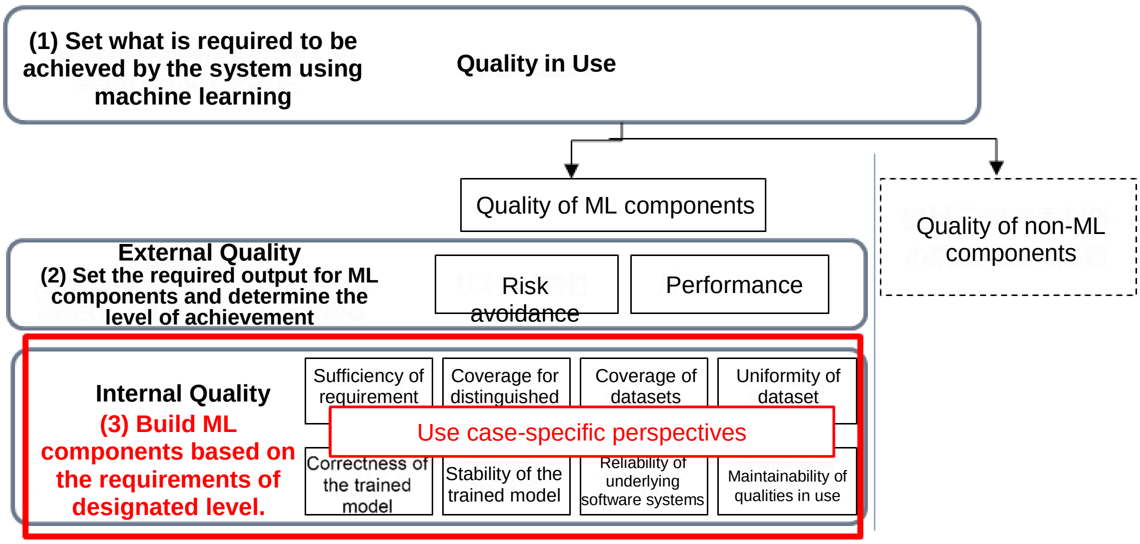
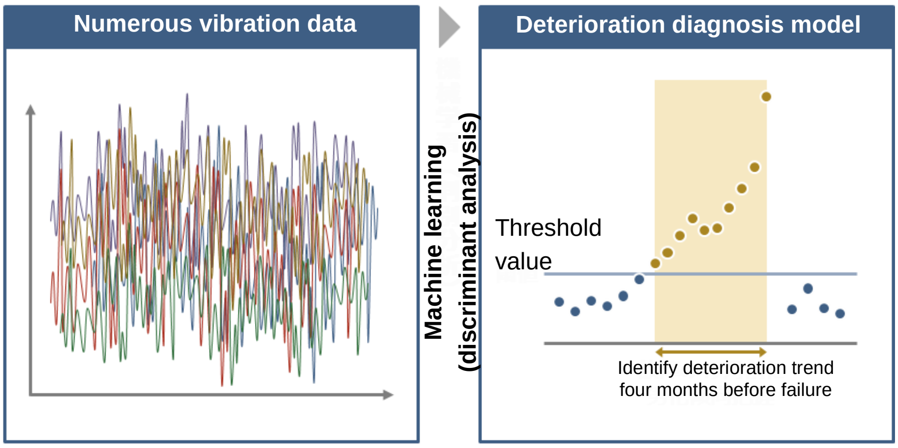


Figure 3-9 Positioning of "Use case-specific perspectives" in hierarchical quality assurance

3.3.3 Equipment deterioration diagnosis

*This case study is provided as a conceptual image of "Equipment deterioration diagnosis" and is not a premise for description in the subsequent use case.

Related case study: "Deterioration diagnosis of agitator" (Plant owner company Y)



* Introduced for the purpose of optimizing maintenance planning and procurement through early identification of deterioration trends (independent of safety-related systems)

Traditionally, vibration sensors installed in various parts of the agitator[71] were used to detect abnormal vibrations several weeks before failure, and decisions were made to stop the equipment and perform maintenance.
However, by analyzing the vibration data and other data using machine learning, it was found that the normal correlation between the bearings of the agitator and the vibration displacement of the reduction gear deviated several months before the failure. If it becomes possible to detect signs of deterioration several months in advance, it is expected that maintenance planning and procurement can be optimized.

The correlation between the vibration displacement of bearings and reduction gears is converted into variables based on past vibration data. Then, the degree of deviation from the normal correlation is indexed by discriminant analysis using Mahalanobis distance[72] and output as the degree of danger. This system is used only for optimizing maintenance and procurement plans in the span of several months, and safety assurance is conducted separately by a system independent of this system.

[71] A device that mixes liquid and liquid, liquid and solid, etc. in a tank in order to promote chemical reactions, mixing, etc.
[72] A quantity that expresses how far a given piece of data is from a data group, taking data variance into account. Often used to determine outlier.

1) Premise of the use case

a. Overview

The use case "equipment deterioration diagnosis" is an ML based system detecting signs of abnormality of a plant equipment that will become apparent in the long-term future (weeks to months), with the aim of quickly detecting deterioration trends in specific parts of plant equipment. If the parts deterioration of plant equipment can be identified at an early stage, plant maintenance will be highly productive, such as reflecting the deterioration in maintenance and procurement plans, and implementing operations to delay the deterioration.
See "3.3.4 Detection and diagnosis of early signs of abnormality" for the use case of the ML based system for detecting abnormality in the plant that may occur in the short-term future, from 20–30 minutes to a few days.

b. Functional requirements

In this use case, "Predicting the future deterioration trends of individual equipment parts" is set as a functional requirement of the ML based system.

c. Conceptual image of introduction

In this use case, a conceptual image of introduction is set as shown in Figure 3-10. In the conventional deterioration diagnosis of plant equipment, a maintenance engineer periodically read the values of multiple sensors installed in the equipment, and when the sensor values deviated from the predetermined thresholds, the decision to replace the equipment was made. On the other hand, after introducing the ML based system, the ML components judge whether or not the equipment has deteriorated, and if judged so, an alert is sent to the maintenance engineer. Maintenance engineers will judge whether the parts need to be inspected or replaced based on the information and the equipment operation data.

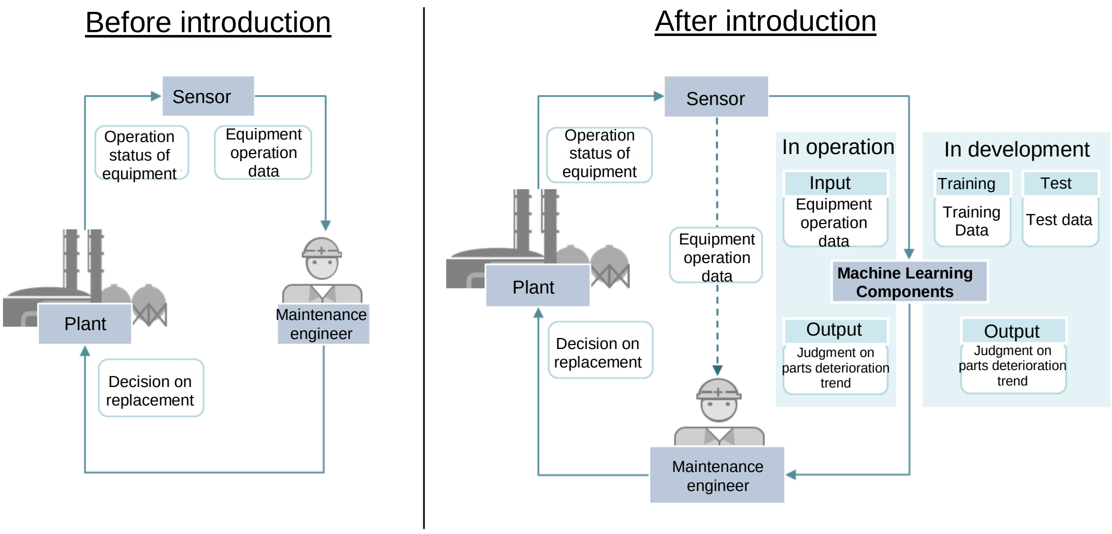


Figure 3-10 Conceptual image (example): Introduction of use case “equipment deterioration diagnosis”

d. Relationship with other systems

In this use case, the relationship between the ML components and other systems is set as shown in Figure 3-11. ML components classify the presence or absence of deterioration trends using equipment operation data as input, and output the results to the maintenance engineer. The maintenance engineer judges whether the deterioration trends of the target equipment are at a level that requires replacement or maintenance, and makes a final decision on whether replacement or maintenance is required. Therefore, under the premise of this use case, the row (3) will be applied when referring to the "AISL Table."
It is assumed that there is no external safety mechanism that monitors and corrects the judgment (presence or absence of deterioration trends of equipment) of ML components.
The ML based system in this use case is designed to diagnose equipment deterioration in order to improve the efficiency of maintenance activities. It does this by detecting signs of failures over the medium to long term, and it is assumed that the existing system for detecting impending equipment failures exists as a completely different system.
In this way, the ML based system is assumed to be purely for the purpose of improving the efficiency of maintenance activities, and safety is assumed to be guaranteed by the existing system. Therefore, under the premise of this use case, “risk avoidance” shall not be set. This is an assessment based on the assumptions illustrated in the Guidelines, and does not mean that "risk avoidance" does not need to be considered in all cases similar to this use case. The level of safety functions required for ML components must be determined for each individual application.

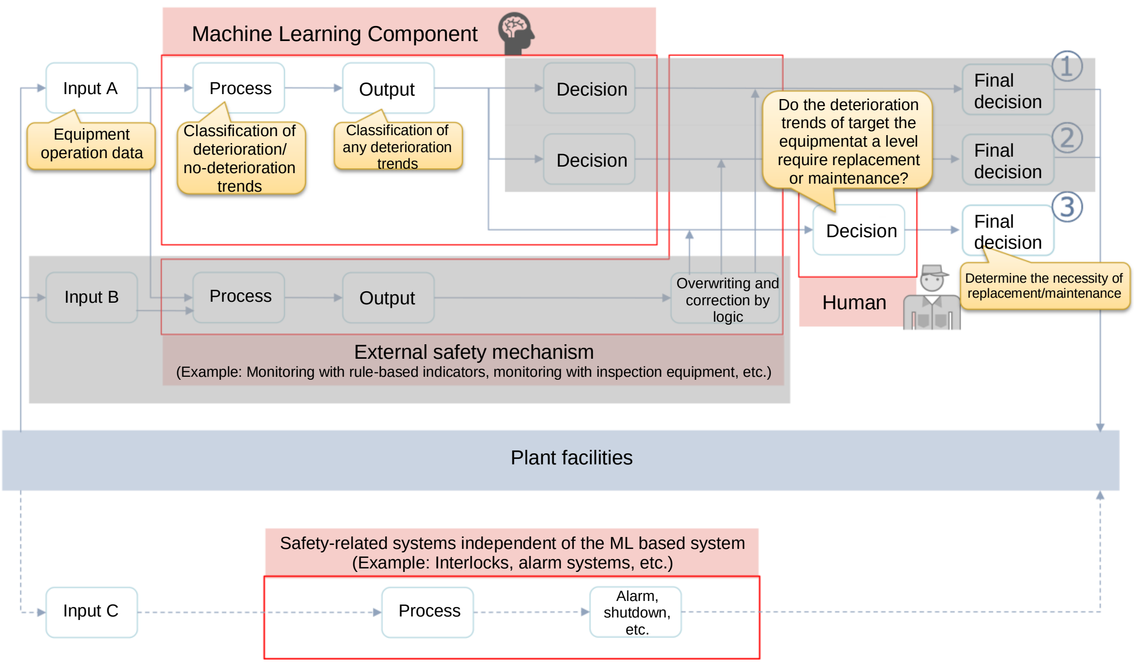


Figure 3-11 Example: Relationship between ML components and other systems in the use case “equipment deterioration diagnosis”

e. Composition of ML components

In this use case, the composition of ML components is set as shown in Table 3-8. As the learning model, a classification model of discriminant analysis that learns the distribution of each variable with and without deterioration is adopted. Therefore, when the chronological operation data of equipment are input at a certain point in time, the trend of detereoration will be output if the operation data are classified into the trained distribution of detereoration. Training data consist of the following: chronological operation data from the past, and a label indicating whether deterioration followed. Test data are constructed in a structure identical to the training data. It is assumed that the label of the existence is given not because the detereoration occuredat the time of data acquisition but it is recorded it would occurr in the future. The "future" here is assumed to be a long-term future of a few weeks to a few months.

Table 3-8 Example: Composition of ML components for “equipment deterioration diagnosis”

Learning methodology	Classification (supervised)
Learning model	Learning the distribution of variables from operation data during deterioration and non-deterioration
Input data in operation	Chronological operation data of equipment
Training data in development	Past chronological operation data of equipment + labels for future deterioration
Test data in development	Past chronological operation data of equipment + labels for future deterioration

2) Examples of items for quality in use and external quality

Assuming that this use case is set, quality in use and external quality items are set according to Table 3-9. In this use case, the ML based system shall be purely for the purpose of improving the efficiency of maintenance activities, and safety is assumed to be guaranteed by the existing system. For this reason, the quality in use and external quality of "risk avoidance" are not set, and the AISL assessment is not conducted. These correspond to “(1) Set what is required to be achieved by the system using machine learning” and “(2) Set the required output for ML components and determine the level of achievement” in the hierarchical quality assurance procedure shown in Chapter 2 (cf. Figure 3-12).

Table 3-9 Example: Quality in use and external quality items for the use case “equipment deterioration diagnosis”

Quality in Use	External Quality
Risk avoidance	
-	-
Performance	
Correctly diagnose the deterioration of parts (P-U1)	Keep the classification error between "deterioration" and "no deterioration" to a certain level (P-E1)
Predict deterioration trends early enough to be reflected in maintenance plans (P-U2)	Output the result of judging the change from "no deterioration" to "deterioration" before a predetermined time (P-E2)

Note: Codes for each item are given for convenience of explanation to clarify the relationship between items, and are not required to be given as the Guidelines, nor are they related to the level of the item. (S: Safety, P: Performance, U: Use, and E: External)

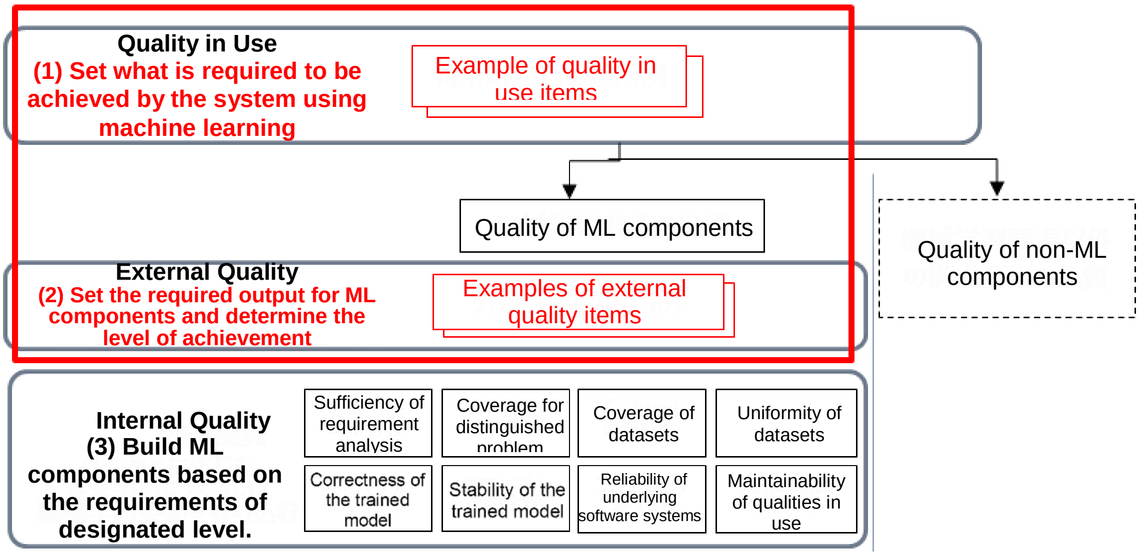


Figure 3-12 Positioning of quality in use/external quality items in hierarchical quality assurance

a. Example of "risk avoidance" Quality

Since the quality in use and external quality of "risk avoidance" are not set in this use case, the AISL assessment is not conducted.

b. Example of "performance" Quality

In light of functional requirement "Predicting the future deterioration trends of individual equipment parts," in order to control quality to achieve functional requirements at a desirable level from the perspective of "Performance," quality in use and external quality shall be specifically set.

- Quality in Use

  In equipment deterioration diagnosis, the accuracy of deterioration diagnosis and the timing of detecting deterioration trends must be set to an appropriate level. Set "Correctly diagnose the deterioration of parts (P-U1)"[73] as the quality in use that defines the accuracy level of deterioration diagnosis. As for the timing of detecting deterioration trends, it is necessary to detect future deterioration trends at a timing that can be reflected in the planning of maintenance in weeks or months. Therefore, "Predict deterioration trends early enough to be reflected in maintenance plans (P-U2)"[74] shall be set as quality in use.

---

[73] In the practical example of this guideline 3. Equipment deterioration diagnosis (Yokogawa Electric Corporation)", quality is set as follows. Consultation is advised.
Quality in use "The state of deterioration should be estimated correctly"
External Quality "The classification errors between "deterioration" and "no deterioration" are kept to a certain level"

[74] In the practical example of this guideline 3. Equipment deterioration diagnosis (Yokogawa Electric Corporation)", quality is set as follows. Consultation is advised.
Quality at Use "The progress of deterioration is judged early enough to develop a maintenance plan."
External Quality "The result of judging the progress of deterioration is output at least X weeks before the time when maintenance is required."

- External Quality

  For each quality in use, the external quality required for the output of ML components is defined. For external quality corresponding to the quality in use, "Correctly diagnose the deterioration of parts (P-U1)," "Keep the classification error between ‘deterioration’ and ‘no deterioration’ to a certain level (P-E1)."[66] In addition, for external quality corresponding to the quality in use, "Predict deterioration trends early enough to be reflected in maintenance plans (P-U2)," "Output the result of judging the change from ‘no deterioration’ to ‘deterioration’ before a predetermined time (P-E2)" is set.

  For each external quality, set the required external quality level "AIPL." For "Keep the classification error between ‘deterioration’ and ‘no deterioration’ to a certain level (P-E1),"[67] consider the requirement level to "Keep the classification error within the target value" and the degree to which it is essential, and set AIPL according to the criteria. Similarly, for "Output the result of judging the change from ‘no deterioration’ to ‘deterioration’ before a predetermined time (P-E2)," consider the level of output timing required and the degree to which it is essential, and set the AIPL according to the criteria. The largest of these AIPLs becomes the AIPL of ML components, and the required level of internal quality is determined accordingly.

3) “Use case-specific perspectives” to ensure internal quality

Based on the premises of this use case, points to consider ("perspective") for fulfilling the requirements of each internal quality are shown in Table 3-10. The “perspectives” described below can be used as a reference when developing ML components similar to this use case. These requirements correspond to “(3) Build ML components based on the requirements of designated level” in the hierarchical quality assurance procedure shown in Chapter 2 (cf. Figure 3-13).

Table 3-10 “Use case-specific perspectives” pertaining to “equipment deterioration diagnosis”

* See the Checklist in Appendix for the list of internal quality requirements and perspectives for this case

Internal Quality	Requirement[75]	Use case-specific perspectives
Sufficiency of requirement analysis	(Common requirements)	・ Determine the range of component values of products to be considered that vary with processing conditions. This includes not only the case where the product to be processed is different, but also the case when the fluid* and the process changes. *Changes in the distribution of mixed flow/multi-phase flows, etc.

[75] In the table, only the requirements related to "Use case-specific perspectives" are excerpted from the "1st edition of Guidelines on Quality Control for Machine Learning." Items that are not listed in this table are also included in the requirements

Internal Quality	Requirement[75]	Use case-specific perspectives
Coverage for distinguished problem cases	・ (Lv1) Furthermore, extract attributes of differences in particularly-important environmental factors and prepare cases corresponding to combinations with serious risk factors.	・ "Environmental factors" in this context refer to location, operating environment, temperature and humidity, operating method, raw materials, utilities, etc.
	(Common requirements)	・ Consider whether the training data can be collected for a range of component values for the target product. ・ When using simulation data, check whether the simulator takes into account changes in environmental factors (e.g. high humidity to low humidity). ・ If a dataset is to be obtained by simulation, the validity of a simulator should be fully verified. ・ The data immediately after a replacement of parts may be "No deterioration." Note that the "No deterioration" period depends on the specifications of the parts and materials, but varies depending on the usage environment (determine the period of "No deterioration" by referring to the frequency of past replacements, etc.). ・ When "running-in[76]" is required immediately after maintenance of the equipment, ensure that no data are collected during that period.
Coverage of datasets	・ (Lv1) Consider the sources and methods of obtaining test data sets so that they are expected to be unbiased in the application status.	・ "Application status" in this context refers to the type and operation status (e.g. constant/temporary, load change) of the target equipment, and others.
	(Common requirements)	・ Personnel with expertise who can make appropriate judgments confirm whether the labels of detereoration are correct.
Uniformity of datasets	(Common requirements)	・ Obtain operation data in various states without bias assumed as "No deterioration."

[76] An operation to check if there are any problems in the condition of equipment by operating at low load immediately after maintenance of the equipment and before full-scale production.

Internal Quality	Requirement[75]	Use case-specific perspectives
		▪ Recognize that if sufficient operation data for a certain state cannot be obtained, the accuracy of detecting deterioration deviating from that state may be reduced.
Correctness of the trained model	-	-
Stability of the trained model	-	-
Reliability of underlying software systems	-	-
Maintainability of qualities in use	(Common requirements)	▪ Note that when the type of the equipment is replaced, it may be necessary to take measures such as re-learning and switching of the learning model. ▪ Ensure that quality is maintained when entering product component values of changing product component values. ▪ Check if there are any deviations from assumptions such as environmental factors that were assumed at the beginning, including not only the target equipment itself but also surrounding conditions.

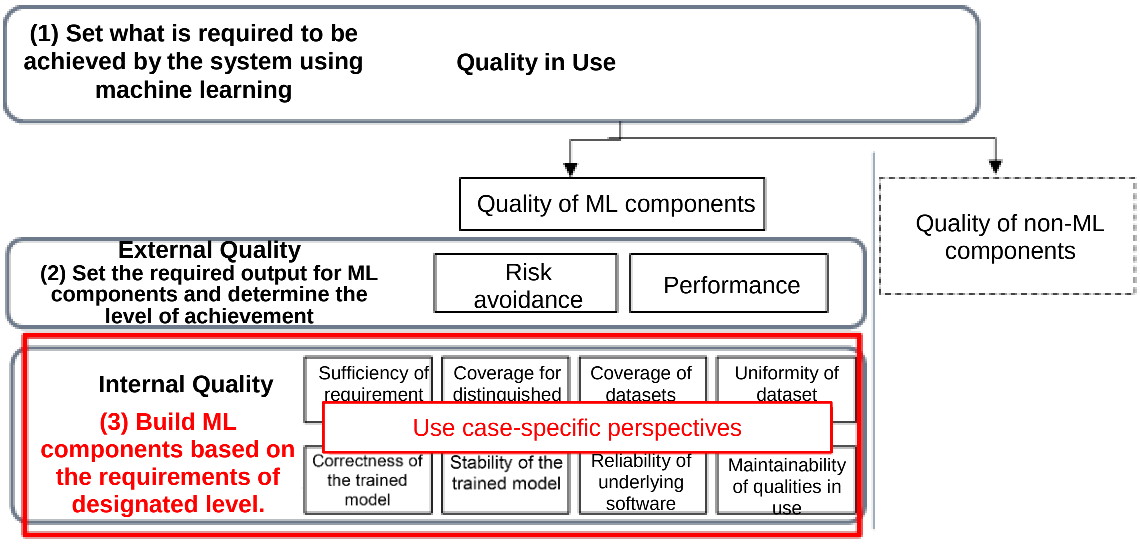
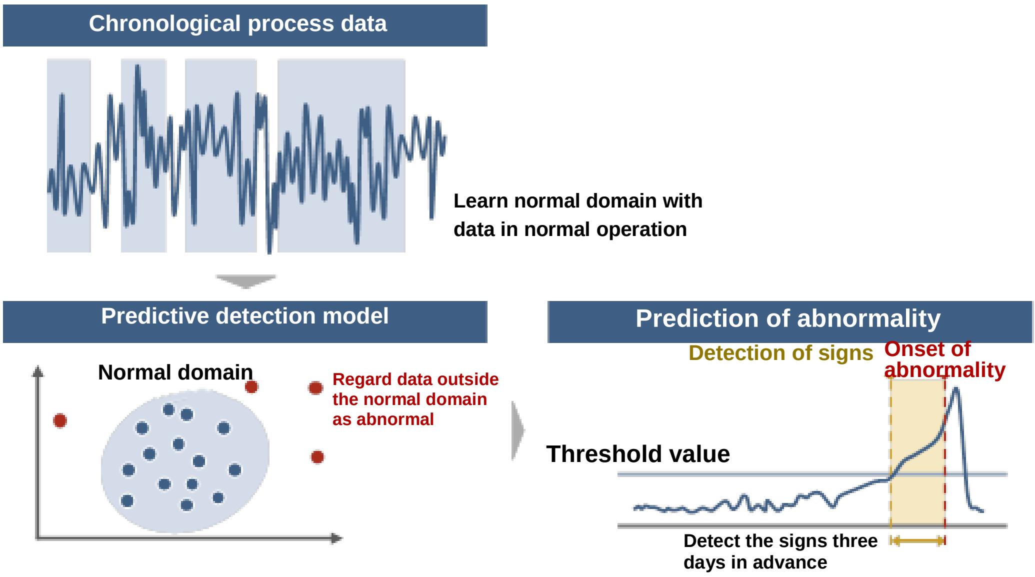


Figure 3-13　Positioning of “Use case-specific perspectives” in hierarchical quality assurance

3.3.4 Detection and diagnosis of early signs of abnormality

*This case study is provided as a conceptual image of "Detection and diagnosis of early signs of abnormality" and is not a premise for description in the subsequent use case.

Related case study: “Detection and diagnosis of early signs of abnormality in power plants” (Plant owner company Z)



The power unit always monitors the upper and lower limits of the operation values of each piece of equipment to ensure that there are no supply disruptions. But there are times when equipment stops unexpectedly due to an abnormality that cannot be detected by upper and lower limit monitoring. Therefore, this system detects the early signs of abnormality to reduce the number of unexpected shutdowns.

An outlier detection model (unsupervised) that uses chronological data in normal time is adopted. Using process operation data obtained from DCS as input, the system displays a graph of the degree of abnormality which quantifies the magnitude of deviation from the normal time and the variables that affect the degree of abnormality.

1) Premise of the use case

a. Overview

The use case "detection and diagnosis of early signs of abnormality" is an ML based system that detects signs of abnormality in advance, which may become apparent in the short-term future, from 20–30 minutes to a few days, in order to avoid unexpected plant shutdowns due to abnormality. When an unexpected shutdown occurs, even if it does not lead to an accident, it incurs a large cost to restart the plant. For this reason, by detecting the signs of abnormality in advance, allowing the plant to stop via a normal procedure and acting on the area where the abnormality has occurred are effective in terms of both safety and productivity. Please refer to "3.3.3 Equipment deterioration diagnosis" for the case of ML components that analyzes the long-term deterioration status of equipment.

b. Functional requirements

In this use case, "Alert when abnormality is detected," which corresponds to the detection of signs of abnormality, and “Output the location and severity of, and the variables related to the abnormality," which corresponds to the diagnosis of the details of abnormality, are set as the functional requirements of the ML based system.

c. Conceptual image of introduction

In this use case, a conceptual image of introduction is set as shown in Figure 3-14. The conventional way of detecting an abnormality at plants involved the operator reading the values of multiple sensors installed in the plant several times a day, and judging whether or not an abnormality would occur in the future based on the operator's experience. On the other hand, after introducing the ML based system, the ML components will monitor in real time whether or not an abnormality will occur in the future (whether or not there are signs of abnormality), and output an alert to the operator if there are signs of abnormality. The operator decides whether to perform certain operations (including plant suspension) based on the output, process data, and others.

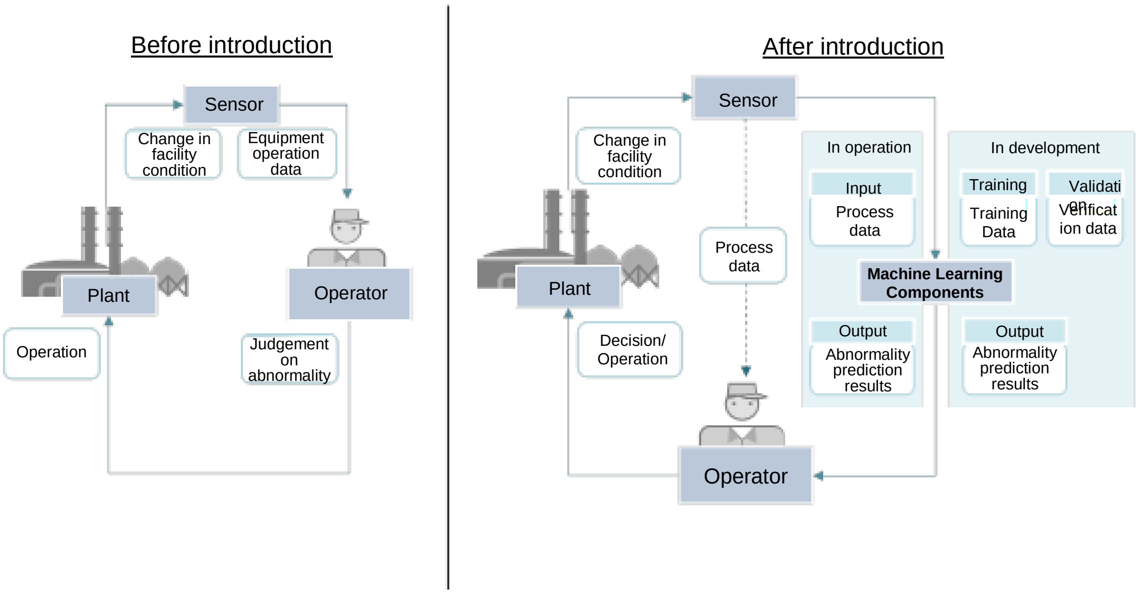


Figure 3-14 Conceptual image (example): Introduction of use case “detection and diagnosis of early signs of abnormality”

d. Relationship with other systems

In this use case, the relationship between the ML components and other systems is set as shown in Figure 3-15. ML components judge whether or not there is a sign of abnormality using process data as input, and alert the operator when it is judged that there is a sign of abnormality. The operator, referring to the judgment of ML components, confirms the data, including the data conventionally monitored by various sensors, etc., and makes a decision to change the operation status, including shutting down the plant. It is assumed that there is no external safety mechanism that monitors and corrects the decision of ML components (any existence of abnormality in the near future), and Safety-related systems independent of the ML based system (e.g. interlocks) exist to ensure plant safety.

Therefore, under the premise of this use case, column (2) will be applied when referring to the "AISL Table."

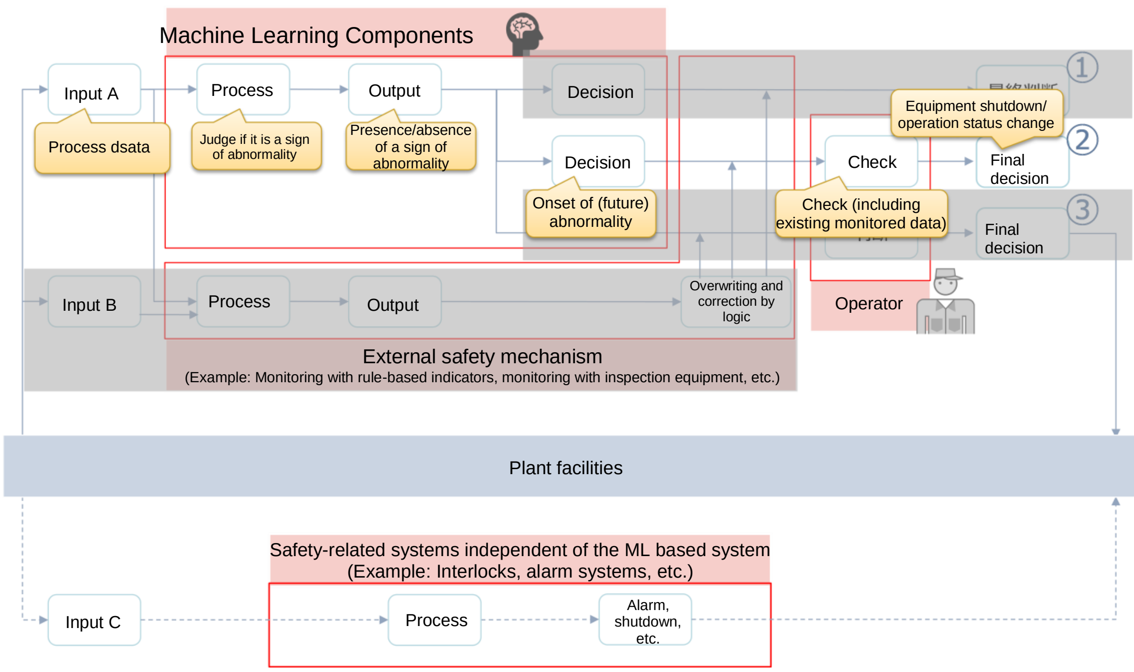


Figure 3-15 Example: Relationship between ML components and other systems in the use case “detection and diagnosis of early signs of abnormality”

e. Composition of ML components

In this use case, the composition of ML components is set as shown in Table 3-11. Because an abnormality occurs infrequently in plants, and it is difficult to collect a large number and various types of data on abnormal incidents, an unsupervised classification model is assumed as the learning model that uses plant data in normal time to learn the normal domain. Therefore, inputs that do not belong to the normal domain are detected as signs of abnormality. It is assumed that the actual process data in normal time is used as training data, and the process data in normal time (actual values) and during an abnormality are used as test data. As the data during an abnormality cannot be tested sufficiently using only the actual measured values, it is assumed that data generated by simulation is used in addition to the actual measured values.

Table 3-11 Example: Composition of ML components for “detection and diagnosis of early signs of abnormality”

Learning methodology	Classification (unsupervised)
Learning model	Learn the normal domain
Input data in operation	Equipment process data
Training data in development	Equipment process data in normal time (actual measured values)
Test data in development	Equipment process data in normal time (actual measured values) Equipment process data during an abnormality (created by actual measured values + simulation)

2) Examples of items for quality in use and external quality

Assuming that this use case is set, quality in use and external quality items are set according to Table 3-12. These correspond to “(1) Set what is required to be achieved by the system using machine learning” and “(2) Set the required output for ML components and determine the level of achievement” in the hierarchical quality assurance procedure shown in Chapter 2 (cf. Figure 3-16).

Table 3-12 Example: Quality in use and external quality items for the use case “detection and diagnosis of early signs of abnormality”

Quality in Use	External Quality
Risk avoidance	
Correctly detect the occurrence of future abnormalities under a variety of plant conditions (S-U1)	In the case of "signs of abnormality," reduce the false-negative rate as much as possible where it is judged to be "normal" (S-E1)
Correctly output the location/severity of and variables correlated with the abnormality under the various plant conditions (S-U2)	In the case of "signs of abnormality," minimize the false recognition rate of location/severity of and variables correlated with the abnormality (S-E2)
The timing of alert should be sufficiently early so that it is possible to take measures to avoid accidents after receiving the alert (S-U3)	Detect by a predetermined time (S-E3)
Performance	
Set the alarm frequency to a reasonable level so that the operators and inspectors do not have to allocate extensive time resource to check the contents of the alarm (P-U1)	Reduce the frequency of false positives below a certain level (P-E1)

Note: Codes for each item are given for convenience of explanation to clarify the relationship between items, and are not required to be given as the Guidelines, nor are they related to the level of the item. (S: Safety, P: Performance, U: Use, and E: External)

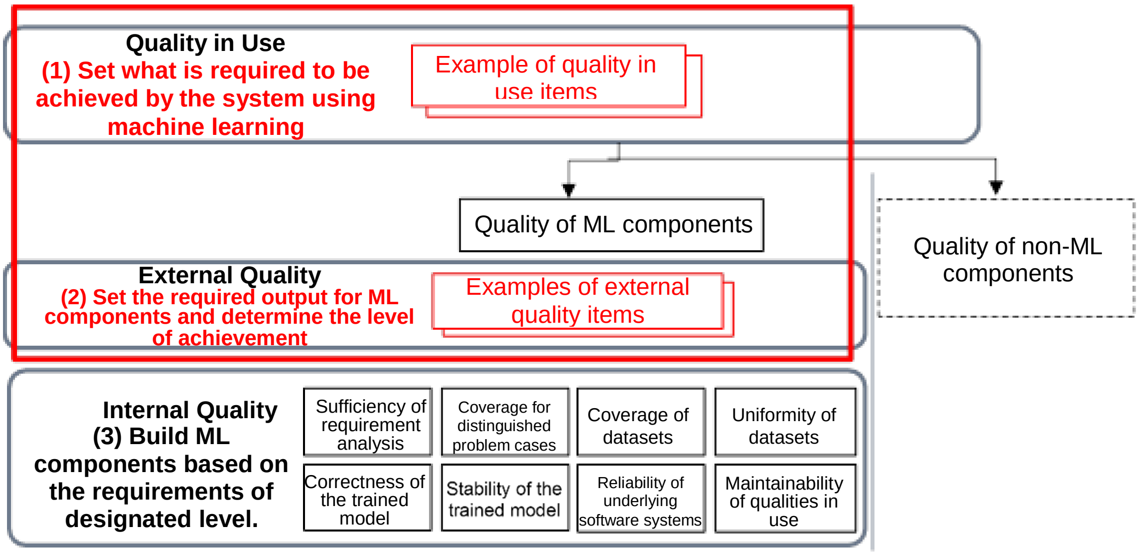


Figure 3-16 Positioning of quality in use/external quality items in hierarchical quality assurance

a. Example of "risk avoidance" Quality

In light of the functional requirements, "Alert at the onset of an abnormality" and "Output the location and severity of, and the variables related to the abnormality," in order to control quality to prevent human and economic damage due to non-achievement of the functional requirement from the "Risk avoidance" perspective, the quality in use and external quality shall be specifically set.

- Quality in Use

  From the perspective of preventing human and economic damage, it is undesirable to miss an abnormality or to output the contents of an abnormality (location, etc.) incorrectly. In order to minimize these risks, set "Correctly detect the occurrence of future abnormality under a variety of plant conditions (S-U1)" and "Correctly output the location/severity of and variables correlated with abnormality under a variety of plant conditions (S-U2)" as quality in use of the Risk Avoidance attribute.[77]." As for the timing of the alert, it is difficult to respond to any alert just before the actual onset of an abnormality (e.g. a few seconds before), so an alert at a sufficiently early timing is required to be meaningful as a predictive detection. In other words, "The timing of alert should be sufficiently early so that it is possible to take measures to avoid accidents after receiving the alert (S-U3)" shall be set as quality in use[78].

[77] In the practical example of this guideline "4-2. Prediction and diagnosis of abnormality (JGC Japan Corporation)", quality is set as follows. Please also refer to it.
Quality in use "The variables that have affected the prediction value are suggested so that the cause of the abnormality prediction can be identified."
External Quality "To minimize errors that erroneously outputs variables that affected a prediction value."

[78] If quality is prioritized to prevent productivity from being reduced due to premature alerting, the timing of alert shall be set as the quality for Performance axis. Risk Avoidance and Performance should be selected individually based on whether the quality to be achieved is related to safety or to efficiency and productivity.

[78] If quality is prioritized to prevent productivity from being reduced due to premature alerting, the timing of alert shall be set as the quality for Performance axis. Risk Avoidance and Performance should be selected individually based on whether the quality to be achieved is related to safety or to efficiency and productivity.

- External Quality

  For each quality in use, the external quality required for the output of ML components is defined. For external quality corresponding to the quality in use "Correctly detect the occurrence of future abnormalities under the various plant conditions (S-U1)," “Minimize the false-negative rate of judging "normal" when "signs of abnormality" are present’ (S-E1)” is set. The reason why the phrase "reduce the false-negative rate as much as possible" is used here instead of "make the false positive/negative rate below a certain level" is, given that the amount and type of past abnormality data that can be used for testing is limited, it is expected that at least the actual values of the limited past abnormality data and the test data that simulates a clearly abnormal situation will be correctly judged as an abnormality.

  For external quality corresponding to the quality in use "Correctly output the location/severity of and variables correlated with the abnormality under the various plant conditions (S-U2)," “In the case of "signs of abnormality," minimize the false recognition rate of location/severity of and variables correlated with the abnormality (S-E2)”[77] is set. The "relation" recognized by the operator is replaced by the output of the ML components and is expressed as a "correlation." In addition, for external quality corresponding to the quality in use "The timing of alert should be sufficiently early so that it is possible to take measures to avoid accidents after receiving the alert (S-U3)," "Detect by a predetermined time (S-E3)" is set.

  For each external quality, set the required external quality level "AISL." For “In the case of "signs of abnormality," reduce the false-negative rate as much as possible where it is judged to be ‘normal’ (S-E1),” consider the magnitude of human and economic damage that could be expected if the abnormality is missed, and set the AISL according to the criteria. For “In the case of ‘signs of abnormality,’ minimize the false recognition rate of location/severity of and variables correlated with the abnormality (S-E2),” consider the magnitude of human and economic damage that could be expected in the event of erroneous output, and set the AISL according to the criteria. Regarding “Detect by a predetermined time (S-E3),” consider the magnitude of human and economic damage that can be expected in the event of a delayed alert, and set the AISL according to the criteria.

  The largest of these AISLs becomes the AISL of ML components, and the required level of internal quality is determined accordingly.

b. Example of "performance" Quality

In light of the functional requirements, "Alert at the onset of an abnormality” and “Output the location and severity of, and the variables related to the abnormality," in order to control the quality to achieve the functional requirements at the desired level from “Performance” perspective, the quality in use and external quality should be specifically set.

- Quality in Use

  In the detection and diagnosis of early signs of abnormality, the accuracy and timing of the alert must be set to a desirable level. As for the accuracy of alerts, a certain amount of false alerts (misrecognition of normal as abnormality) can be tolerated. But if the frequency of false alerts is too high, the confirmation work and plant shutdown time will increase, and this will have a significant adverse effect on plant operations, which will be intolerable. Therefore, “Set the alarm frequency to a reasonable level that does not require long hours for operators and inspectors to check the contents of the alarm (P-U1)" shall be set as quality in use.

- External Quality

  For each quality in use, the external quality required for the output of ML components is defined. For external quality corresponding to the quality in use "Set the alarm frequency to a reasonable level that does not require long hours for operators and inspectors to check the contents of the alarm (P-U1)," "Reduce the frequency of false positives below a certain level (P-E1)" is set. The "reasonable alert frequency" recognized by the operator is replaced by the output of the ML components, and is expressed as "Reduce the frequency of false positives below a certain level."

  For this external quality, set the required external quality level "AIPL." For "Reduce the frequency of false positives below a certain level (P-E1)," the level of the required accuracy rate and the degree to which it is essential are examined, and the AIPL is set according to the criteria. The set AIPL becomes the AIPL of ML components, and the required level of internal quality is determined accordingly.

3) "Use case-specific perspectives" to ensure internal quality

Based on the premises of this use case, points to consider ("perspective") for fulfilling the requirements of each internal quality are shown in Table 3-13. The "perspectives" described below can be used as a reference when developing ML components similar to this use case. These requirements correspond to "(3) Build ML components based on the requirements of designated level" in the hierarchical quality assurance procedure shown in Chapter 2 (cf. Figure 3-5).

Table 3-13 "Use case-specific perspectives" pertaining to "detection and diagnosis of early signs of abnormality"

* See the Checklist in Appendix for the list of internal quality requirements and perspectives for this case

Internal Quality	Requirement[79]	Use case-specific perspectives
Sufficiency of requirement analysis	(Common requirements)	・ Since the evaluation of "Coverage for distinguished problem cases" and "Coverage of datasets" is affected by types and location of targeted abnormality, specify the requirements including types and location. ・ Even if a causal relationship in an engineering sense between the detection of an abnormality and the related variables is not accounted for, it is acceptable to use the suggested relationship for inferrence.[80]

[79] In the table, only the requirements related to "Use case-specific perspectives" are excerpted from the "1st edition of Guidelines on Quality Control for Machine Learning." Items that are not listed in this table are also included in the requirements.

[80] In the detection and diagnosis of early signs of abnormality, some argue that "It cannot be used unless the cause of

Internal Quality	Requirement[79]	Use case-specific perspectives
Coverage for distinguished problem cases	・ (Lv1) Furthermore, extract attributes of differences in particularly-important environmental factors and prepare cases corresponding to combinations with serious risk factors.	・ "Environmental factors" in this context refer to factors which affect the detection of abnormality (e.g. production load, production lot).
	・ (Common requirements)	・ If a data set is to be obtained by simulation, the validity of a simulator should be fully verified.
Coverage of datasets	・ (Lv1) Consider the sources and methods of obtaining test data sets so that they are expected to be unbiased in the application status.	・ "Application status" in this context refers to the severity of an abnormality to be detected and the situation of use of an ML based system (e.g. regular/temporary, daytime/night, steady/unsteady).
	・ (Lv1) Ensure that no bias is expected by performing unbiased sample extraction from original data for each case.	・ In this case, it is not mandatory to cover the data of all casesof abnormality as training data. On the other hand, exhaustive sample extraction in normal domain is required.
	(Common requirements)	・ Personnel with expertise who can make appropriate judgments confirm that the data under normal conditions is actually data under such conditions.
Uniformity of datasets	(Common requirements)	・ Obtain data without bias in various ranges (e.g. daytime/night, steady/unsteady, seasonal differences) assumed as normal data. ・ Recognize that if sufficient normal data for a certain range cannot be obtained, the accuracy of detecting abnormality within that range may be reduced.

abnormality is identified and there is an engineering account of the causal relationship between the cause and abnormality." On the other hand, there is also an aspect where the usefulness of machine learning is to find a correlation with the causal relationship unknown. The Guidelines take the position that, from the perspective of promoting the use of machine learning, the use of correlations alone is not precluded, even if a meaningful engineering causality is unknown, on the premise that safety is ensured by having an agreement between users and vendors.

Internal Quality	Requirement[79]	Use case-specific perspectives
Correctness of the trained model	・ (Lv1) When allowing a certain number of misjudgments during the test phase (including the case of changing the treatment with false positive/negative), reasonably determine and record the criteria in advance.	・ In this case, a certain amount of false detection can be tolerated, but since the amount and types of abnormal data that can be used for the test are limited, it is preferable for false detection rate to be as close to zero as possible.
Stability of the trained model	-	-
Reliability of underlying software systems	-	-
Maintainability of qualities in use	(Common requirements)	・ Since changes in the external environment (e.g. sunlight, wind direction) have a particularly large impact on ML components in chemical plants, pay attention to changes that affect the external environment of the target equipment, even if they are not changes of the equipment, such as removal or modification of adjacent equipment. ・ Expect aging to progress depending on the production load of the target equipment and design the frequency of accuracy verification and tuning of the learning model accordingly. ・ Accuracy verification and tuning of the learning model are necessary each time when the target equipment is repaired on a large scale (not aging).

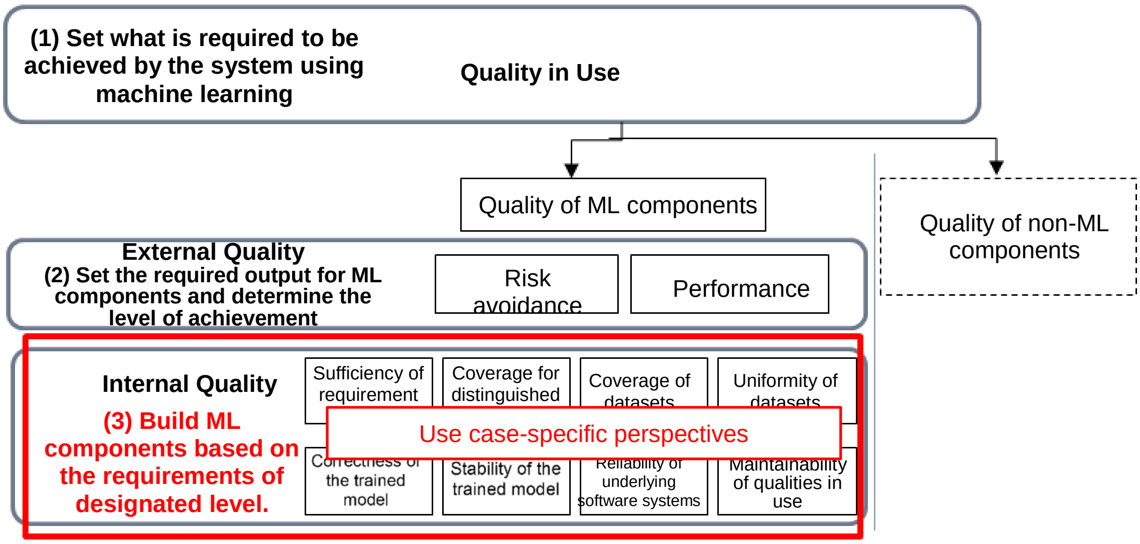
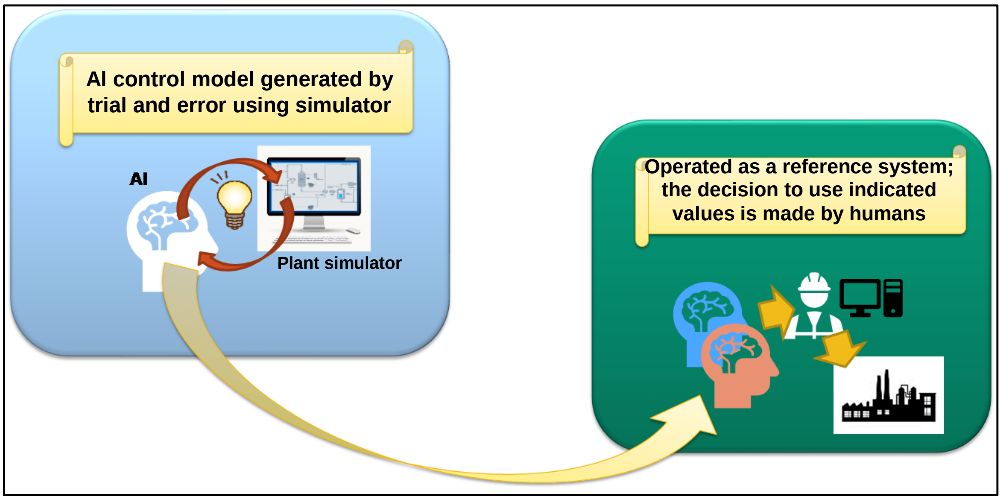


Figure 3-17 Positioning of “Use case-specific perspectives” in hierarchical quality assurance

3.3.5 Optimization of operation

*This case study is provided as a conceptual image of "optimization of operation" and is not a premise for description in the subsequent use case.

Related case study (steady operation): Construction of AI control system for butadiene production plant
(Yokogawa Electric Corporation)

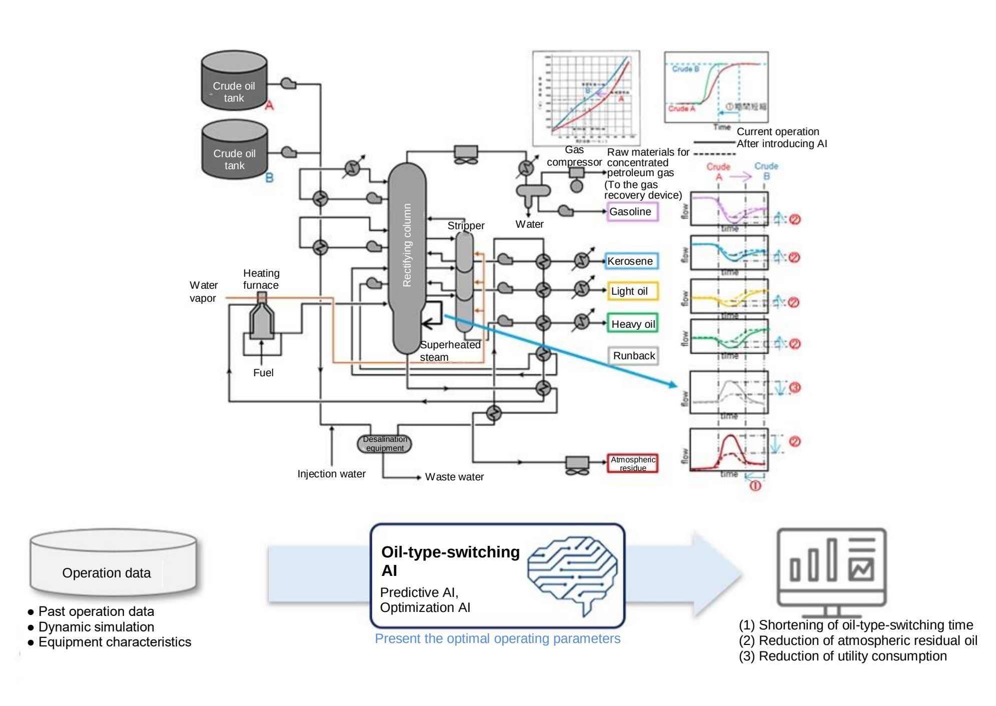


In butadiene refining plants, PID control and multivariable model predictive control are used to automate and stabilize plant operations. However, it is difficult to fully automate some of the processes that are greatly affected by external disturbances such as weather changes, so the plant is currently being controlled by manual operation by operators and other methods.

To resolve this, a simulator is built for the plant to train AI automatically using reinforcement learning, and the model obtained from the training is applied to the actual plant. In this way, automation of the target process is expanded to a range that is difficult to achieve with existing control methods, human errors are reduced, and productivity and operation accuracy are improved.

*The case study is provided as a conceptual image of "optimization of operation" and is not a premise for description in the subsequent use case.

Related case study (unsteady operation): AI for optimization of crude oil switching operation at refineries
(Chiyoda Corporation)[81]



In a petroleum refinery plant, every time a tank of imported crude oil is nearing empty, a switching operation is performed. Here, an experienced operator continuously adjusts more than ten parameters simultaneously according to the difference in properties to the next oil type. This operation requires about half a day of manual operation once every three days or so, which is both frequent and difficult, and a wrong operation can lead to equipment damage and hinder operations. Therefore, if optimal operation can be achieved, the effect of improving productivity and safety could be significant.
By using a dynamic plant simulator and deep reinforcement learning, AI can output the optimal operating parameters. The optimal operating parameters are presented to the operator in real time during the oil-type switching operation to save energy, minimize product loss, quickly complete the switching, and ensure safe operation.

Deep reinforcement learning utilizes the optimization targets (energy conservation, minimization of product loss, etc.) for a given operating conditions. These optimization targets are built from the knowledge of experienced operators and past operation data. The dynamic plant simulator and AI will work together to learn the operating parameters that will be highly scored for various operating conditions, so that they can continuously find better operating parameters even under new conditions.

[81] This case study is described in detail in the "Collection of Case Examples of Leading Companies Introducing AI into Plants - Practical Examples of Achieving Results and Breaking Through Challenges in AI Projects -." Ensure you read it as a reference.

1) Premise of the use case

a. Overview

The use case "optimization of operation" is an ML based system that presents the optimal operating parameters for the purpose of achieving optimization goals of the plant, such as maximizing product productivity, minimizing product loss, and saving energy. Plant operations can be broadly divided into steady operations and unsteady operations. Unsteady operations include plant startup/shutdown, product changeover, and sudden shutdown, while steady operations include other relatively stable operations. This use case assumes a function to present the optimal operating parameters to the operator based on the operating status of each plant facility, production data, weather, raw materials, and various conditions, which are common to both steady and unsteady operations. For the following use case explanation that apply only to either steady or unsteady operations, whether the intended operation is steady or unsteady is indicated.

b. Functional requirements

In this use case, "Present the optimal operating parameters according to the purpose" is set as a functional requirement of the ML based system. The purpose of optimization is assumed to be to improve productivity in case of steady operations and have an early transition to steady operations in case of unsteady operations.

c. Conceptual image of introduction

In this use case, a conceptual image of introduction is set as shown in Figure 3-18. In conventional plant operation, an operator grasped the state of equipment and environment using data from various sensors installed in the plant. He/she then adjusted the operating parameters according to the operation goals based on past experience. On the other hand, after introducing the ML based system, ML components calculate the optimal operating parameters in real time from the plant equipment and environmental data, which are then presented to the operator. Operators run the plant by referring to the presented operating parameters, using sensor data and their own experience.

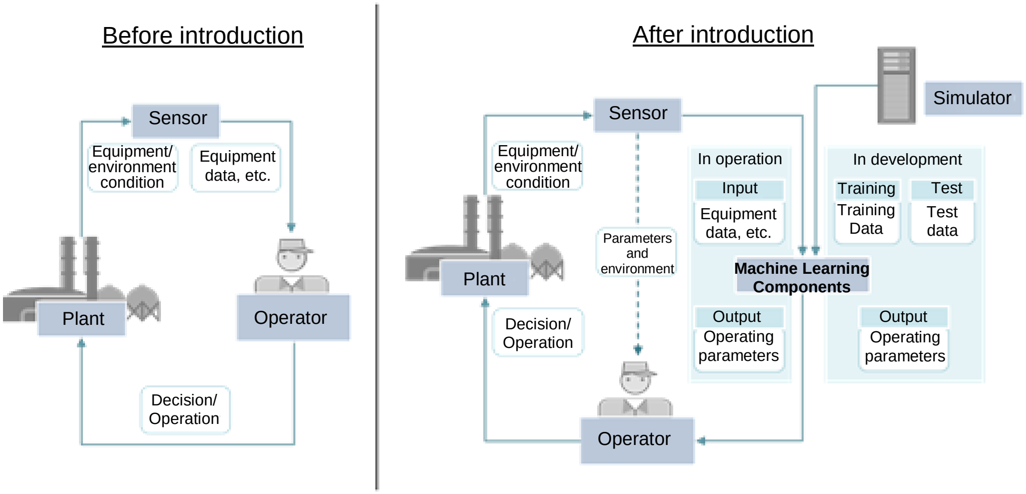


Figure 3-18　Conceptual image (example): Introduction of use case “optimization of operation”

d. Relationship with other systems

In this use case, the relationship between the ML components and other systems is set as shown in Figure 3-19. ML components input various pieces of sensor data of the plant (equipment data, environmental data, production data, etc.) Equipment data assume the number of equipment rotations, temperature, etc.; environmental data assume temperature, humidity, etc.; and production data assume production volume, etc. Optimal values of operating parameters are calculated from these inputs according to the optimization target, and output to the operator. The operator, referring to his/her own experience and various pieces of sensor data, decides whether the operating parameters output by the ML components are appropriate or not, and decides which parameters to be actually used for operating the plant. Therefore, under the premise of this use case, column (3) will be applied when referring to the "AISL Table."
As an external safety mechanism that monitors and corrects the output of ML components (optimal values of operating parameters), it is assumed that there is a process that determines whether the current equipment condition has reached a risk level and monitors the operating parameters output by ML components. It is also assumed that existing systems such as the conventional alert system and emergency shutdown system also exist in the plant independently.

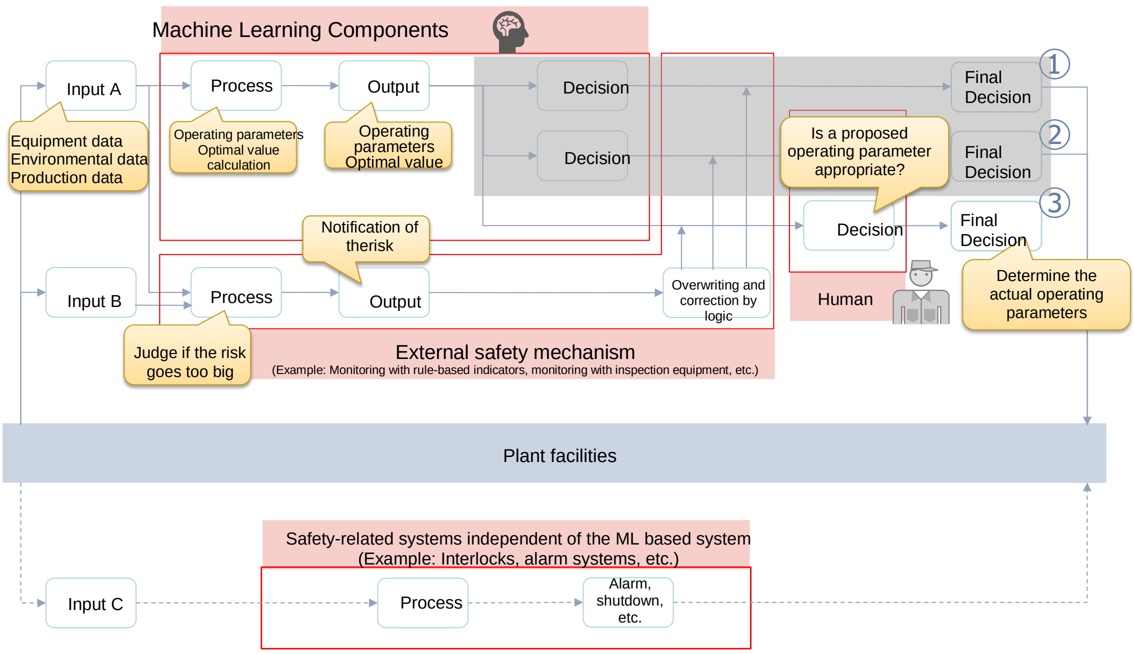


Figure 3-19 Example: Relationship between ML components and other systems in the use case "optimization of operation"

e. Composition of ML components

In this use case, the composition of ML components is set as shown in Table 3-14. As a learning model, the reinforcement learning model is assumed that learns optimal operating parameters. Reinforcement learning assumes the use of a simulator because it is difficult to satisfy the coverage and sufficiency of learning and test data only with actual measurements at the plant, such as operation near the upper limit of equipment parameters.

Table 3-14　Example: Composition of ML components for “optimization of operation”

Learning methodology	Reinforcement learning
Learning model	Reinforcement learning for optimal operating parameters
Input data in operation	Equipment data, environmental data, production data
Training data in development	Equipment data, environmental data, production data + operating parameters (created by actual measurements + simulation)
Test data in development	Equipment data, environmental data, production data + operating parameters (created by actual measurements + simulation)

2) Examples of items for quality in use and external quality

Assuming that this use case is set, quality in use and external quality items are set according to Table 3-15. These correspond to “(1) Set what is required to be achieved by the system using machine learning” and “(2) Set the required output for ML components and determine the level of achievement” in the hierarchical quality assurance procedure shown in Chapter 2 (cf. Figure 3-20).

Table 3-15　Example: Quality in use and external quality items for the use case “optimization of operation”

Quality in Use	External Quality
Risk avoidance	
Do not cause operating conditions that exceed the allowable safety operating specifications of the equipment (S-U1)	Limit the range of parameter variables of the equipment to be optimized to the range corresponding to allowable safety operating specifications (S-E1)
Performance	
Steady operation: Provide parameters that improve productivity (P-U1-1)	Steady operation: Present parameters that improve production by a certain percentage (P-E1-1)
Unsteady operation: Provide parameters for early transition to the steady operation, etc. (P-U1-2)	Unsteady operation: Provide parameters to shorten the transition time to steady operation by a certain percentage, etc. (P-E1-2)

Note: Codes for each item are given for convenience of explanation to clarify the relationship between items, and are not required to be given as the Guidelines, nor are they related to the level of the item. (S: Safety, P: Performance, U: Use, and E: External)

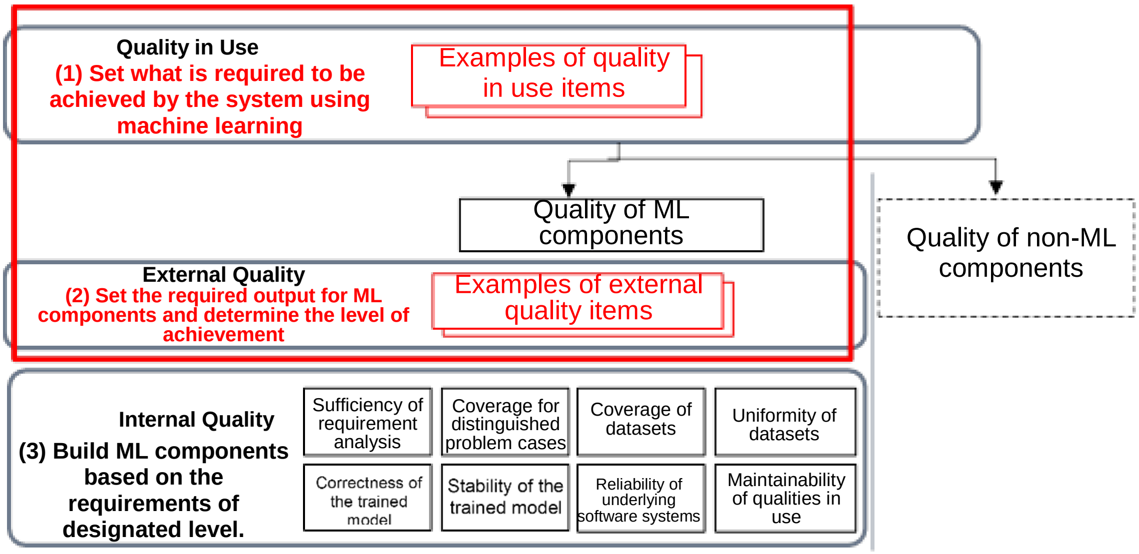


Figure 3-20 Positioning of quality in use/external quality items in hierarchical quality assurance

a. Example of "risk avoidance" Quality

In light of the functional requirement "Present the optimal operating parameters according to the purpose," in order to control quality to prevent human and economic damage due to non-achievement of the functional requirements from the perspective of "risk avoidance," the quality in use and external quality to be controlled shall be specifically set.

- Quality in Use

  From the perspective of preventing human and economic damage, it is not desirable for the output operating parameters to exceed the assumed equipment specifications of the equipment. In order to minimize this risk, the quality in use of the "risk avoidance" attribute should be set to "Do not cause operating conditions that exceed the allowable safety operating specifications of the equipment (S-U1)."

- External Quality

  For the set quality in use, seek the external quality required for the output of ML components. For external quality corresponding to the quality in use "Do not cause operating conditions that exceed the allowable safety operating specifications of the equipment (S-U1)," "Limit the range of parameter variables of the equipment to be optimized to the range corresponding to allowable safety operating specifications (S-E1)" is set.
  For this external quality, the required external quality level "AISL" is set. AISL is set according to the standard by considering the external safety mechanisms, Safety-related systems independent of the ML based system, and examining the possible effects when operating parameters that exceed the assumed equipment specifications are presented.
  In this use case, it is assumed that external safety mechanisms with secured reliability, Safety-related systems independent of the ML based system (e.g. emergency shutdown system) ensure safety. Therefore, considering that there is no need to require a high level of safety for ML components, an SIL assessment is performed and "no SIL" is assigned. Then it is considered

that, based on the AISL Table, AISL from 0 to 0.2 is assigned. This is a judgment based on the assumptions illustrated in the Guidelines, and it does not mean that the assessment will be the same for all cases similar to this use case. The level of safety functions required for ML components must be determined for each individual application.

b. Example of "performance" Quality

In light of functional requirement "Present the optimal operating parameters according to the purpose," in order to control quality to achieve functional requirements at a desirable level from the perspective of "Performance," quality in use and external quality to be controlled shall be specifically set.

- Quality in Use

  In optimizing operations, it is necessary to achieve the optimization goals such as improvement of productivity (steady operation) and early transition to the steady operation state (unsteady operation). For this reason, quality in use is set to "Provide parameters that improve productivity (P-U1-1)"[82][83] and "Provide parameters for early transition to steady operation, etc. (P-U1-2)."

- External Quality

  For the set quality in use, determine the external quality required for the output of ML components. In the case of steady operation, "Present parameters that improve production by a certain percentage (P-E1-1)"[74] is set for the external quality corresponding to the quality in use "Provide parameters that improve productivity (P-U1-1)." In the case of unsteady operation, "Provide parameters to shorten the transition time to steady operation by a certain percentage (P-E1-2)" is set for the external quality corresponding to the quality in use "Provide parameters for early transition to the steady operation (P-U1-2)."
  For each external quality, set the required external quality level "AIPL." In the case of steady operation, for "Provide parameters to increase production by a certain percentage (P-E1-1)," the level of the required production volume and the degree to which it is essential are considered, and the AIPL is set according to the criteria. Similarly, for "Provide parameters for early transition to steady operation (P-E1-2)," the level of reduction in transition time and the degree to which it is essential are considered, and the AIPL is set according to the criteria. The required level of internal quality is determined accordingly.

---

[82] Under Practical Example "5-1. Optimization of operation (ENEOS Corporation and Preferred Networks, Inc.)" in the Guidelines, quality is set as follows. Please also refer to it.
Quality in use "To realize optimum control in consideration of both production efficiency and energy efficiency when conditions are steady."
External Quality "To indicate an operation parameter that keeps an indicator, obtained by converting production efficiency and energy efficiency into a monetary value, higher than a certain level."

[83] Under Practical Example "5-2. Optimization of operation (Yokogawa Electric Corporation, JSR Corporation)" in the Guidelines, quality is set as follows. Please also refer to it.
Quality at Use "More energy-saving operation than before is achieved."
External Quality "Operating parameters that keep energy saving indicators above a certain level are presented.

"[84] In the table, only the requirements related to "Use case-specific perspectives" are excerpted from the "1st edition of Guidelines on Quality Control for Machine Learning." Items that are not listed in this table are also included in the requirements.

3) “Use case-specific perspectives” to ensure internal quality

Based on the premises of this use case, points to consider ("perspective") for fulfilling the requirements of each internal quality are shown in Table 3-16. The “perspectives” described below can be used as a reference when developing ML components similar to this use case. These requirements correspond to “(3) Build ML components based on the requirements of designated level” in the hierarchical quality assurance procedure shown in Chapter 2 (cf. Figure 3-5).

Table 3-16 “Use case-specific perspectives” pertaining to “optimization of operation”

* See the Checklist in Appendix for the list of internal quality requirements and perspectives for this case

<table>
<tr><th rowspan="2">Internal Quality</th><th rowspan="2">Requirement[84]</th><th colspan="2">Use case-specific perspectives</th></tr>
<tr><th>Steady operation</th><th>Unsteady operation</th></tr>
<tr><td>Sufficiency of requirement analysis</td><td>-</td><td colspan="2">-</td></tr>
<tr><td rowspan="2">Coverage for distinguished problem cases</td><td>・ (Lv1) Furthermore, extract attributes of differences in particularly-important environmental factors and prepare cases corresponding to combinations with serious risk factors.</td><td colspan="2">・ "Environmental factors" in this context refer to operating procedures, raw materials, and others.</td></tr>
<tr><td>・ (Common requirements)</td><td colspan="2">・ If a data set is to be obtained by simulation, the validity of a simulator should be fully verified.</td></tr>
<tr><td rowspan="2">Coverage of datasets</td><td>・ (Lv1) Consider the sources and methods of obtaining test data sets so that they are expected to be unbiased in the application status.</td><td>・ "Application status" in this context refers to the applicable operation situation (season, time of day, etc.) and the equipment to be operated.</td><td>・ "Application status" in this context refers to the applicable operation situation (season, time of day, start-up/shutdown, etc.) and the equipment to be operated.</td></tr>
<tr><td>・ (Common requirements)</td><td colspan="2">・ Pay attention to whether the assumed data set includes "disturbance" such as weather.
・ When learning the operations of experienced operators, make sure that the case setting is not biased.</td></tr>
<tr><td>Uniformity of datasets</td><td>-</td><td colspan="2">-</td></tr>
</table>

[84] In the table, only the requirements related to "Use case-specific perspectives" are excerpted from the "1st edition of Guidelines on Quality Control for Machine Learning." Items that are not listed in this table are also included in the requirements.

<table>
<tr><th rowspan="2">Internal Quality</th><th rowspan="2">Requirement[84]</th><th colspan="2">Use case-specific perspectives</th></tr>
<tr><th>Steady operation</th><th>Unsteady operation</th></tr>
<tr><td rowspan="2">Correctness of the trained model</td><td>・ Decide and record how to deal with the incorrect behavior of a trained model (e.g., false negative/false positive in the test) before the validation phase.</td><td colspan="2">・ In the case of "Assumption that no safety function is required from the presentation of operating parameters by ML components (when operating parameters are presented that exceed the assumed specifications of the equipment, they are not reflected in the operation by external safety mechanisms and operator judgment)," it is not required to consider "Keeping the permissible level of output of parameters leading to dangerous operation as close to zero as possible." Based on the SIL assessment, etc., determine the required level for the ML components by appropriately assigningthe safety functions to the ML components and other systems.</td></tr>
<tr><td>・ (Common requirements)</td><td colspan="2">・ Even if reinforcement learning is used, meet the requirements of "Correctness of the trained model" by conducting tests before the start of operation. [85]</td></tr>
<tr><td>Stability of the trained model</td><td>-</td><td colspan="2">-</td></tr>
<tr><td>Reliability of underlying software systems</td><td>-</td><td colspan="2">-</td></tr>
<tr><td>Maintainability of qualities in use</td><td>(Common requirements)</td><td colspan="2">・ Monitor the evaluation of optimal values on a regular and continuous basis to check for abnormalities.
・ When a change is made to equipment or operating procedures, update the model because it will affect the output of ML components.
・ As stability may be impaired if operating conditions are pursued to the limit with respect to the optimization target, take measures such as limiting the output range of ML components.
・ Confirm that the equipment is operated within the expected</td></tr>
</table>

[85] In the case of reinforcement learning, it is assumed that the system will be put into operation without testing and higher performance will be pursued while the system is in operation, and in this case, the "Correctness of the trained model" requirement cannot be applied. However, the field of plant safety requires testing to confirm that a certain level of performance has been achieved at the start of operation. Hence, the Guidelines require that testing be conducted before the start of operation to satisfy the "Correctness of the trained model" requirement even if reinforcement learning is used.

Internal Quality	Requirement[84]	Use case-specific perspectives	
		Steady operation	Unsteady operation
		interpolation range of the assumed raw material (e.g. crude oil type). · Confirm the output quality of ML components by considering various conditions (e.g. initial period of/end of reaction, operating conditions, raw materials, quality requirements, allowable time for startup and shutdown) of equipment under operation.	

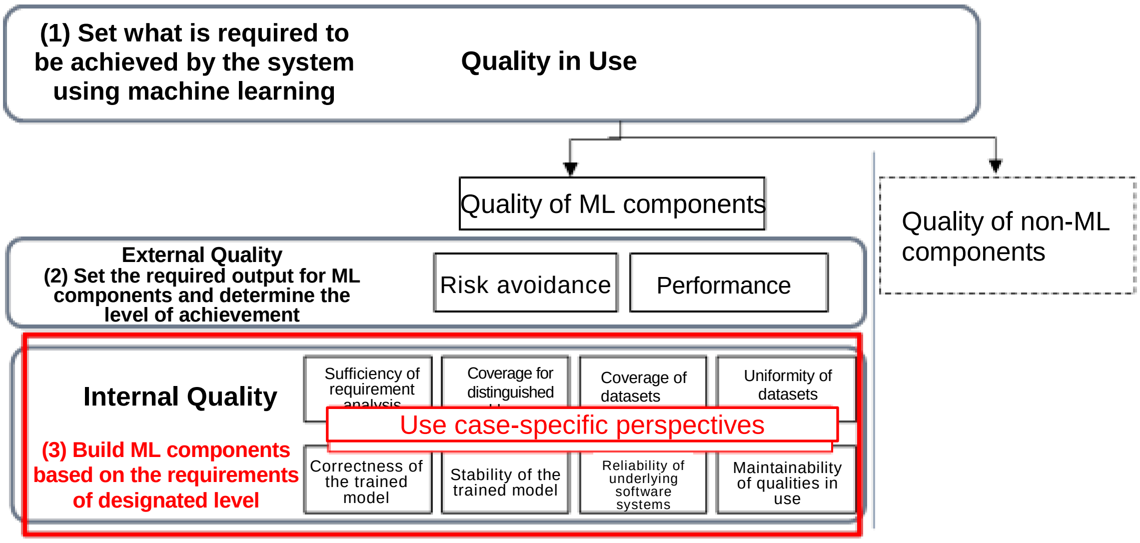


Figure 3-21 Positioning of "Use case-specific perspectives" in hierarchical quality assurance

4. Flow of utilizing the Guidelines

Chapter 2 provided the specific method of assessing the reliability of machine learning in the field of plant safety, and Chapter 3 introduced examples based on the use case. This chapter will show how to apply Chapters 2 and 3 in accordance with the concrete steps of reliability assessment.

First, regarding the key players in the application of Guidelines, their roles and for what purposes business providers can refer to the Guidelines are organized (4.1). In system development and operation, a concrete flow of applying the Guidelines on reliability assessment, such as the specific roles assumed by each key player in reliability assessment (maintenance staff, plant system staff, etc.), and the section of the Guidelines to be referred, is descrebed(4.2).

4.1 Key players in utilizing the Guidelines

The reliability assessment based on the Guidelines is conducted in each process from requirements definition to operation of the ML based system. In this process, various players from multiple companies are involved. Table 4-1 shows the types of players involved, the affiliated companies, and the roles generally required in the field of plant safety. Each staff member should be involved in ensuring the reliability of the ML based system from their respective positions.

Table 4-1　Roles in the development project of the ML based system

Department/Role	Description of Role in the Project
Corporate Planning	Decision-making on project execution *Non-key player in utilizing the Guidelines
Project Planning staff	Project Lead of the ML based system development (budget, schedule, etc.)
Quality Assurance staff (ML based system, ML component)	Assess and confirm the quality in use of the overall ML based system Assess and confirm the external and internal qualities of the ML based system
Environment & Safety staff	Review the contents related to safety of the ML based systemML based system[86]
Field staff Field manager (Manufacturing staff, Facility Management staff), Field Operation staff (Operator, maintenance engineer)	Review the quality in use in systems used for operation and maintenance Review the external quality required for ML components Review whether the desired outcome is achieved in operation Review the external safety mechanisms and safety-related systems
Plant System staff	Set the quality in use of the ML based system and external quality of ML components Organize non-ML components (external safety mechanism, etc.) of the ML based systemML based system and safety-related systems Review and provide data pertaining to the development of ML components
ML Design & Development staff	Design and development of ML components

It is characteristic of the field of plant safety that the staff members who are directly responsible for operation and maintenance (Table 4-1 "Field staff") review the quality of the overall ML based system and ML components, and the operation management staff of the plant's existing systems including DCS[87] (Table 4-1 "Plant System staff") are involved in providing data and defining requirements, etc. for the ML based system.

The post of "Quality Assurance staff" may need to be assumed by a proper organization or personnel with appropriate level of independence from other staff involved in the development of the ML based system. The level of independence should be determined according to the existing standards (e.g.

[86] "Review" shall mean that, upon receiving a request from the Project Lead engaged in reliability assessments using the Guidelines, the items considered by the Project Lead are confirmed based on respective duties and expertise. For example, if the functional requirements and quality in use of the ML based system are set by the "Project Planning staff," who leads the project, and the "Field staff," who will ultimately use the system, confirm that these requirements are set correctly, the "Project Planning staff" will be the Project Lead and the "Field staff" will be the Reviewer. The Reviewer does not necessarily need to read and understand the contents of the Guidelines, but should be involved in the reliability assessment in response to a request from the Project Lead. See Section 4.2 for detailed implementation.

[87] Distributed Control System (DCS) is a plant control system. Rather than controlling the entire plant with a single control device, there is a control device for each component, and these devices are connected by a network to realize control of large-scale plant operations.

functional safety standards) to which the plant adheres.

For reference, examples[88] of the affiliated company by role are shown in Table 4-2.
It is assumed that the user companies (plant owners, etc.) that consider and decide on whether to introduce the ML based system will take the lead in the development project[89]. In some cases, plant equipment vendor companies may provide data related to the plant equipment, and system vendor companies, SIer companies, and AI vendor companies may participate in the design and development of the ML based system and ML components.

Table 4-2　Example: Company affiliation by role in ML based system development project

Department/Role		Example of Affiliated Company
Corporate Planning		Plant owner company
Project Planning staff		Plant owner company
Quality Assurance staff	Entire system	System vendor/SIer company Plant owner company
	Machine Learning Components	AI vendor company
Environment & Safety staff		Plant owner company
Field staff	Manufacturing staff Facility Management staff	Plant owner company Plant equipment vendor company
	Operator	Plant owner company
	Maintenance engineer	Plant owner company Maintenance company
Plant System staff		System vendor/SIer company Plant owner company
ML Design & Development staff		AI vendor company

Since roles may vary depending on the project, the abovementioned role assignments are only examples for reference. It is important to assign the roles of Key Staff and a Reviewer appropriately to clarify who is involved in setting the quality without fixing the roles to be "Key staff → User company; Reviewer → Vendor company.”

[88] The companies that actually participate in ML based system development projects and the roles of each organization vary depending on the individual project. Please also refer to "Practical Examples" of the Guidelines, where the staff members are listed based on actual reliability assessment cases conducted in the real projects.
[89] In the development project of ML based systems, in addition to the pattern of the user enterprise initiative exemplified in this guideline, there is also a pattern of proceeding based on the proposal of the vendor enterprise while obtaining the confirmation of the user enterprise. Please also refer to "Practical Examples" of the Guidelines, where examples of actual reliability assessment conducted under both user-led and vendor-led patterns are shown. (Practical example led by a vendor company: "1. Prediction of pipe wall thickness (Yokogawa Electric Corporation)", "3. Equipment deterioration diagnosis (Yokogawa Electric Corporation)", "4-1. Prediction and diagnosis of abnormality (Chiyoda Corporation and Seibu Oil Company Limited)", "4-2. Prediction and diagnosis of abnormality (JGC Japan Corporation)")

Since the Guidelines are applied in different situations depending on the role of each staff member, the quality that should be focused on in confirmation varies among the qualities set in the three stages. Qualities to be confirmed by each staff are shown in Table 4-3.
By checking 2.1 and 2.2, which explains the three qualitis, each staff member needs to understand qualities that they are concerned with.

Table 4-3　Application of the Guidelines and quality to be checked by each staff member

Department/ Role	Application of the Guidelines	Quality in Use	External Quality	Internal Quality
Corporate Planning	-	○	-	-
Project Planning staff	Setting the purpose of entire system and quality in use; presenting them to Corporate Planning	● → 2.1.1 2.1.3 2.2.1	○	○
Quality Assurance staff	Confirming the quality in use Confirming the external and internal qualities	● → 2.1.1 2.1.3 2.2.1	● → 2.1.2 2.1.3 2.2.2 2.2.3	● → 2.1.4 2.2.4 2.2.5
Environment & Safety staff		○	○	○
Field staff	Reviewing the quality in use Reviewing the external quality Reviewing the outcome and quality in operation	○	○	○
Plant System staff	Setting the quality in use and external quality; providing/reviewing internal quality data	● → 2.1.1 2.1.3 2.2.1	● → 2.1.2 2.1.3 2.2.2 2.2.3	● → 2.1.4 2.2.4 2.2.5
ML Design & Development staff	Grasping the internal quality; designing, developing and updating ML components in line with the requirements of internal quality	-	○	● → 2.1.4 2.2.4 2.2.5

Legend
● : Key players carrying out the reliability assessment procedure in accordance with the Guidelines
○ : Reviewers of reliability assessment (there is no need to read and understand the contents of the Guidelines, but be involved in the reliability assessment upon request from Project Lead)

4.2 Flow of applying the Guidelines

This section describes how the key players listed in 4.1 assume roles in the reliability assessment and utilize the Guidelines in each phase of the ML based system development (from requirement definition to testing and acceptance) and the operation.

4.2.1 Activity items by role and phase

An overview of activity items related to reliability assessment by role and phase is shown in Table 4-4.

Participating key players vary depending on whether the ML based system is for maintenance or operation. In the case of the ML based system for maintenance, it is assumed that the equipment manager (field manager) and maintenance engineer (field operator) will be involved as field staff. For operation, on the other hand, it is assumed that the manufacturing staff (field manager) and operator and maintenance engineer (field operator) will be involved. Therefore, in this chapter and the "Appendix: “Perspectives in the Field of Plant Safety” Checklist for Internal Quality Assurance" they are distinguished.

Table 4-4 Activity items by phase and role for the development and operation of the ML based system

Role		PoC	Requirement definition	Design	Implementation	Testing & acceptance	Operation
Corporate Planning			Make decision on system development			Make decision on system implementation	
Project Planning staff		(Considered by core development members, referring to the implementations in each phase on the right)	Set the purpose of the system; review the system functional requirements and quality in use			Review the overall quality in acceptance inspection	
Quality assurance	System Quality Assurance staff					Assess the **external quality** based on testing; acceptance inspection of the system	Confirm the **quality in use and external quality**
	ML Quality Assurance staff				Quality assurance in the development of ML components		Confirm the **internal quality**
Environment & Safety staff				Review the external safety mechanisms and safety-related systems pertaining to the external quality		Review the external quality assessment results based on testing	
Field staff	Field manager [Maintenance] Facility Management staff [Operation] Manufacturing staff		Review the purpose of the system, functional requirements and quality in use	Review the external quality and the level setting; review the external safety mechanisms and safety-related systems pertaining to the external quality	Review pertaining to the ML component development	Review the overall quality in acceptance inspection	Review the results of confirmation of quality in use; review and provide data pertaining to system updates
	Field Operation staff [Maintenance] Maintenance engineer [Operation] Operator, maintenance engineer		Review the purpose of the system, functional requirements and quality in use		Review pertaining to the ML component development	Review the overall quality in acceptance inspection	Review the results of confirmation of quality in use; review and provide data pertaining to system updates
Plant System staff			Set the system functional requirements and **quality in use**	Set the **external quality**; organize non-ML components (external safety mechanism, etc.) and safety-related systems pertaining to the level setting of **external quality**; set the level of **external quality**	Develop non-ML components of the ML based system (external safety mechanism, etc.); review and provide data pertaining to the ML component development	Review the external quality assessment results based on testing	Review the results of confirmation of external quality; update non-ML components (external safety mechanism, etc.)
ML Design & Development staff					Set the level of **internal quality**; design and develop the ML components		Update the ML components pertaining to **internal quality**

Legend in **bold and underlined**: Key implementing players of quality assurance activities related to the quality; underlined only: key players that support quality assurance activities related to the relevant quality through review, etc.

Note: Internal quality development are included in the "implementation" phase.

Table 4-5 shows the activity items for the reliability assessment in each phase in the development and operation of the ML based system, shown in the horizontal axis of Table 4-4.

Table 4-5 Activity items by phase for the development and operation of the ML based system

Phase	Step	Activity Item in Quality Assurance
PoC	-	-
Requirement definition	1	Setting the purpose of the system[90]
	2	Functional requirements of the system Setting the quality in use
Design	3	Setting the external quality
	4	Summerizeng the ML based system and safety-related systems pertaining to the level of external quality
	5	Setting the level of external quality
Implementation	6	Setting the level of internal quality
	7	Designing and development of ML components
	8	Developing the non-ML components of the ML based system (external safety mechanism, etc.)
Testing & acceptance	9	Testing (assessment of external quality)
	10	Performing acceptance inspection
Operation	11	Confirming the quality in use
	12	Confirming the external quality
	13	Confirming the internal quality (Maintainability of qualities in use)
	14	Systeming updating

In the Guidelines, the phases are described in chronological order (requirement definition → design → implementation → testing and acceptance → operation), but the Guidelines are not necessarily applicable only to the waterfall style of development.
It may be done iteratively in each phase and between phases, as shown in Figure 4-1. For example, in

[90] In actual AI development, there are many cases where the purpose of the system is not clear at the beginning, but the Guidelines are intended for consideration after the user's issue is determined. Consideration on searching for such issues is a step before the reliability assessment is required. In addition, please refer to the "Collection of Advanced AI Case Examples at the Plants," which describes the issue of "difficulty in objective setting for AI projects" and the solutions to overcome this difficulty.

[91] At the stage of this step "Setting External Quality", it is not necessary to set numerical targets specific to ML (Example: Accuracy greater than or equal to ○%). In addition, regarding the external quality of "risk avoidance," it may be possible to define numerical targets required of ML components (e.g. the incidence of misjudgments leading to danger) by checking safety-related systems and external safety mechanisms in the process of Step 5 Setting the level of external quality. At the stage of building an ML component (Step 7 "Design and development of an ML component" in Chapter 4 of the Guidelines), concrete numerical targets specific to machine learning (e.g. accuracy, F-measure) can be set according to the results of PoC, data acquisition status, learning status, etc.

the testing phase, if it is assessed that the reliability (e.g. to minimize the frequency of false detection as an external quality) specified in the design phase is not sufficiently achieved, it is effective to go back to the previous phase/step, check why the reliability is not sufficiently achieved, and implement it again in order. In this way, it is also possible to conduct reliability assessment by moving back and forth between the implementations in each phase and step as appropriate.
In addition, the Guidelines assume that the design and development of ML components (Step 7) are generally done in an agile style, so applying the Guidelines does not necessarily require a change in the development style of ML components.

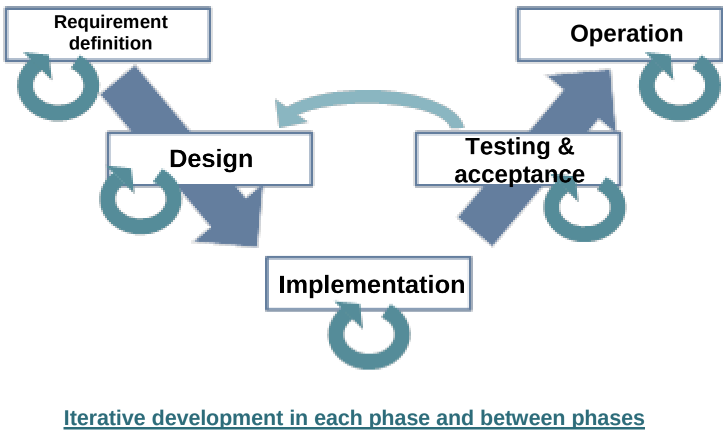


Figure 4-1 System construction and operation style with iterations between phases

4.2.2 Overview of activity items and description

An overview of implementation staff, activity items and description is listed in Table 4-6. Phases and roles described in the following table are typical examples and should be set depending on the project.

Table 4-6 List of activity items and contents for assessing the reliability of machine learning in the field of plant safety

(Legend Implementer in **bold**: Key implementing player for the activity item; Small letters: Reviewer of the activity item)

Phase	Step	Activity Item in Quality Assurance	Implementer	Description	Guidelines
PoC	-	-	-	-	-
Requirement definition	1	Setting the purpose of the system	**Project Planning staff**	Organize the issues to be addressed; set the "purpose of the system"; and judge whether an ML based system is necessary	-
			Field staff	Review the tasks and purpose of the system from the standpoint of the field users	
	2	Functional requirements of the system ・ Setting the quality in use	**Plant System staff**	Set functional requirements and quality in use for the ML based system	2.1.1 2.1.3 2.2.1 (3.3)
			Project Planning staff	Review the functional requirements and quality in use of the set ML based system from the standpoint of Planning staff	
			Field staff	Review the functional requirements and quality in use of the set ML based system from the standpoint of the field and users	
Design	3	Setting the external quality	**Plant System staff**	Set the external quality of ML components based on the quality in use	2.1.2 2.1.3 2.2.2 (3.3)
			Field manager	Review the embodiment of external quality set from the standpoint of the field manager	
	4	Summarizing the ML based system and safety-related systems pertaining to the level of external quality	**Plant System staff**	Organize each component of the ML based system (ML components and non-ML components (external safety mechanism, etc.)), existence of safety-related systems independent of the system, and functional requirements	2.2.3
			Field manager	Review the non-ML components (external safety mechanism, etc.) of the ML based system and independent safety-related systems from the standpoint of the field manager	
			Environment & Safety staff	Review non-ML components (external safety mechanism, etc.) of the ML based system and safety-related systems independent of the system from the standpoint of safety management	

Phase	Step	Activity Item in Quality Assurance	Implementer	Description	Guidelines
	5	Setting the level of external quality	**Plant System staff**	Set the level of external quality of ML components	2.2.3
			Field manager	Review the level of external quality from the standpoint of the field	
Implementation	6	Setting the level of internal quality	**ML Design & Development staff**	Set the level of internal quality of ML components based on the level of external quality	2.1.4 2.2.4 Appendix (Checklist) (3.3)
	7	Designing and development of ML components	**ML Design & Development staff**	Develop according to the design of ML components, internal quality requirements and perspectives	2.2.5 Appendix (Checklist)
			ML Quality Assurance staff	In the development of ML components, check whether the internal quality level and perspectives are satisfied from the standpoint of quality assurance	
			Field staff	Provide data in the development of ML components from the standpoint of the field users	
			Plant System staff	Provide data in the development of ML components from the standpoint of data management	
	8	Developing the non-ML components (external safety mechanism, etc.)	**Plant System staff**	Develop the non-ML components required for external quality and quality in use (external safety mechanism, etc.)	-
Testing & acceptance	9	Testing (assessment of external quality)	**System Quality Assurance staff**	Test the ML based system and assess the results. Assess whether the external quality of ML components meets the required level	2.2.3
			Plant System staff	Based on the assessment of test results, review from the standpoint of having organized external safety mechanisms and safety-related systems and having set the external quality and its level	
			Environment & Safety staff	Review the assessment on test results from the standpoint of safety management	

Phase	Step	Activity Item in Quality Assurance	Implementer	Description	Guidelines
	10	Performing acceptance inspection	**System Quality Assurance staff**	Assess overall ML based system based on the results of assessing external quality in the test, and if the ML based system satisfys the criteria based on the assessment results, etc., conduct an acceptance inspection	-
			Project Planning staff	Review the overall ML based system from the standpoint of project originator	
			Field staff	Review the quality in use assessment results from the standpoint of the field users	
Operation	11	Confirming the quality in use	**System Quality Assurance staff**	Confirm the quality in use of the ML based system in operation	2.1.1 2.1.3 2.2.1
			Field staff	Review the results of confirmation of quality in use from the standpoint of the field users	
	12	Confirming the external quality	**System Quality Assurance staff**	Check the external quality based on the results of confirming quality in use of the ML based system in operation	2.2.3
			Plant System staff	Review the results of confirmation of external quality from the standpoint of having organized external safety mechanisms and safety-related systems and having set the external quality and its level	
	13	Confirming the internal quality (Maintainability of qualities in use)	**ML Quality Assurance staff**	Confirm measures taken to deal with internal quality requirements and perspectives based on the results of confirmation of the quality in use and external quality of the ML based system in operation	2.2.4 2.2.5 Appendix (Checklist)
	14	System updating	**ML Design & Development staff**	Update ML components according to the results of the external quality and internal quality checks	2.2.4 2.2.5 Appendix (Checklist)
			Plant System staff	Update non-ML components (e.g. External safety mechanism) according to the results of confirmation of the quality in use and external quality, and provide data for updating ML components from the standpoint of system management	
			Field staff	Review from the standpoint of the field users in updating ML components	

4.2.3 Activity items and description by phase

(1) PoC

PoC is a concept that covers various meanings, from pure trial study to preparation for full-scale development, and not all cases require consideration of reliability assessment. However, in the later stage of the PoC phase, when full-scale development is expected, it is effective to consider in advance the implementations ofreliability assessment in each phase as a preparation for smooth reliability assessment in the subsequent development. Specifically, the items listed in Table 4-7 should be considered at the end of the PoC phase in order to facilitate the development process. It is difficult to set the exact level of external quality at the PoC phase when the specific system and facilities are not yet determined. However, it is effective to forecast the required level of external quality in light of the functional requirements and quality in use, and to consider the policy for creating internal quality.
For specific implementation details, please refer to the description of the relevant phase.
In full-scale development, each activity item will be carried out by staff with respective roles. But in the PoC phase, it is assumed that those roles will be presumably considered by the core members of the development team (Project Planning staff, Plant System staff, ML Design & Development staff, etc.).

Table 4-7 Items that should be identified at the end of PoC

Phase	Step	Activity Item in Quality Assurance	Relevant Section of the Guidelines
Requirement definition (4.2.3 (2))	1 (4.2.3 (2)1))	Set the purpose of the system	-
	2 ((2)2))	Set the system functional requirements and <u>quality in use</u>	2.1.1/2.1.3/2.2.1 (3.3)
Design (4.2.3 (3))	3 (3.2.1 (3) 3))	Set the <u>external quality</u>	2.1.2/2.1.3/2.2.2 (3.3)
	5 ((3) 5))	Set the level of <u>external quality</u>	2.2.3
Implementation (4.2.3 (4))	6 (4.2.3 (4) 6))	Set the level of <u>internal quality</u>	2.1.4/2.2.4/Appendix (Checklist) (3.3)

Note: The numbers in parentheses in the phase steps are the item numbers in Section 4.2 of the Guidelines where the corresponding phase steps are listed.

(2) Requirement definition

In the requirement definition, plant owner companies first identify issues to be addressed, and then set the functional requirements and quality in use of the system to resolve them.

The decision to use machine learning is made based on the necessity of problem-solving. If ML components are required as a result of the problem identification, reliability assessment shall be conducted based on the Guidelines. If ML components are not required, the conventional reliability assessment is performed as a system that does not include ML components.

Table 4-8 Implementations in the "requirement definition" phase

Step	Activity Item in Quality Assurance	Implementer	Relevant Section of the Guidelines
1	Set the purpose of the system	Project Planning staff: Organize the issues to be addressed by introducing the system, set the purpose of the system (Field staff: Review the issues and purpose of the system from the standpoint of the field users)	-
2	Set the system functional requirements and quality in use	Plant System staff: Set the functional requirements and quality in use for the ML based system (Project Planning staff: Review the functional requirements and quality in use of the ML based system from the standpoint of Planning staff Field staff: Review the functional requirements and quality in use of the ML based system from the standpoint of the field users)	2.1.1/2.1.3/2.2.1: Confirm the defining, axis, and setting method of quality in use (Refer to 3.3 for concrete examples of functional requirements and quality in use)

1) Setting purpose of the system

- Implementations

Project Planning staff

Organize the current issues to be addressed; set the purpose of the system; and judge whether an ML based system is necessary

(For example, if pipeline image diagnosis is to be used to reduce the cost of pipe inspections, consider what specific costs and issues are involved)

Field staff

Review the tasks and purpose from the standpoint of the field users

- Relevant section of the Guidelines

Although this activity is an entrance to quality assurance, it is not included in the Guidelines as it is one step before specific quality assurance and reliability assessment activities.

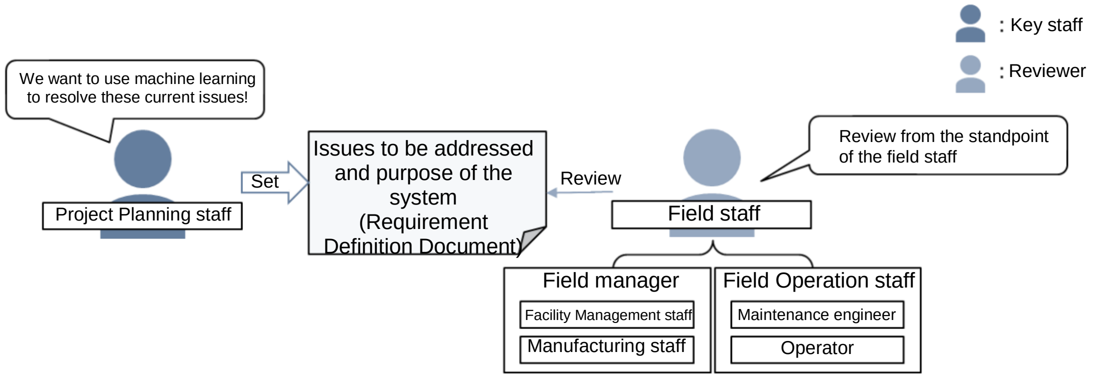


Figure 4-2　Conceptual image of implementation of Step 1 “Setting purpose of the system”

2) Setting system functional requirements and quality in use

- Implementations

Plant System staff

Set functional requirements and quality in use for the ML based system

Project Planning staff

Review the functional requirements and quality in use of the set ML based system from the standpoint of Planning staff

Field staff

Review the functional requirements and quality in use of the set ML based system from the standpoint of the field users

- Relevant section of the Guidelines

Plant System staff will understand the position of quality in use in "2.1.1 Quality in use" and "2.1.3 Quality in use and external quality axes," and how to set quality in use in "2.2.1 Setting quality in use." Set the quality in use of the ML based system according to the described method. Project Planning and Field staff will refer to the same section as needed.
Plant System staff shall refer, as necessary, to "3.3 Specific application of reliability assessment based on use cases" as examples of functional requirements and quality in use for use cases similar to the ML based system to be built this time.

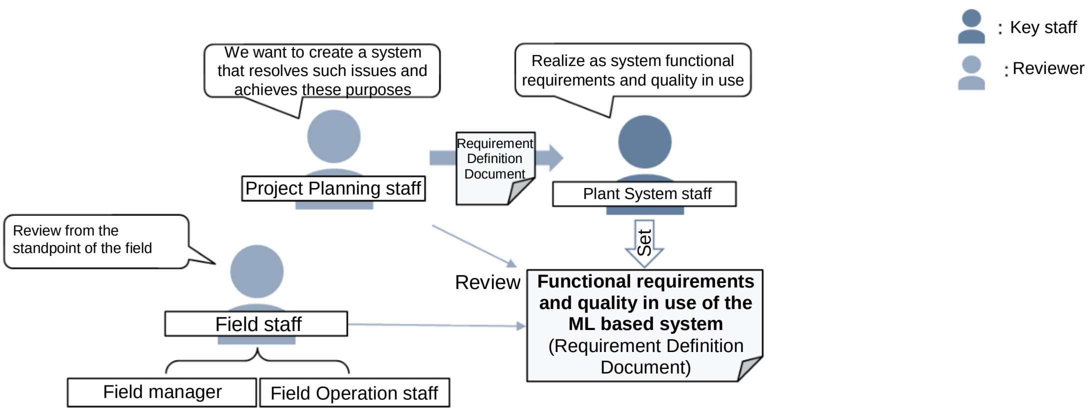


Figure 4-3 Conceptual image of implementation of Step 2 “Setting system functional requirements and quality in use”

(3) Design

Design the overall ML based system. In other words, based on the quality in use required for the overall ML based system, the requirements for each component, such as the external quality of ML components and the roles of the system other than ML components, are established.
In addition, it is important to consider the frequency and implementation criteria of the monitoring required in the "(6) Operation" phase in advance to confirm each quality in the "(6) Operation" phase.

Table 4-9 Implementations in the “design” phase

Step	Activity Item in Quality Assurance	Implementer	Relevant Section of the Guidelines
3	Setting the external quality	Plant System staff: Set the external quality of ML components based on the quality in use (Field manager: Review the embodiment of external quality set from the standpoint of the field manager)	2.1.2/2.1.3/2.2.2: Confirm the positioning and setting method of external quality (Refer to 3.3 as concrete examples of external quality)
4	Organizing the ML based system and safety-related systems pertaining to the level of external quality	Plant System staff: Organize each component of the ML based system (ML components and non-ML components (external safety mechanism, etc.)), existence of safety-related systems independent of the system, relationship and functional requirements Field manager: Review non-ML components (external safety mechanisms, etc.) of the ML based system and safety-related systems independent of the system from the standpoint of the field manager Environment & Safety staff: Review non-ML components (external safety mechanism, etc.) of the ML based system and safety-related systems independent of the system from the standpoint of safety management)	2.2.3: Confirm the method of setting the level of external quality based on the position of non-ML components and independent safety-related systems
5	Setting the level of external quality	Plant System staff: Set the level of external quality of ML components (Field manager: Review the level of external quality from the standpoint of the field manager)	2.2.3: Confirm the method of setting the level of external quality

3) Setting the external quality[91]

- Implementations

Plant System staff

Set the external quality of ML components based on the contents of embodied quality in use

Field manager

Review the set external quality from the standpoint of the field manager

- Relevant section of the Guidelines

Plant System staff needs to understand the definition of external quality in "2.1.2 External quality" and "2.1.3 Quality in use and external quality axes," and how to set external quality in "2.2.2 Setting external quality." The external quality of ML components is set according to the described method. The Field Manager will refer to the same section as needed.
Plant System staff will, as needed, refer to specific examples of external quality of use cases similar to the ML based system to be built this time in "3.3 Specific application of reliability assessment based on use cases."

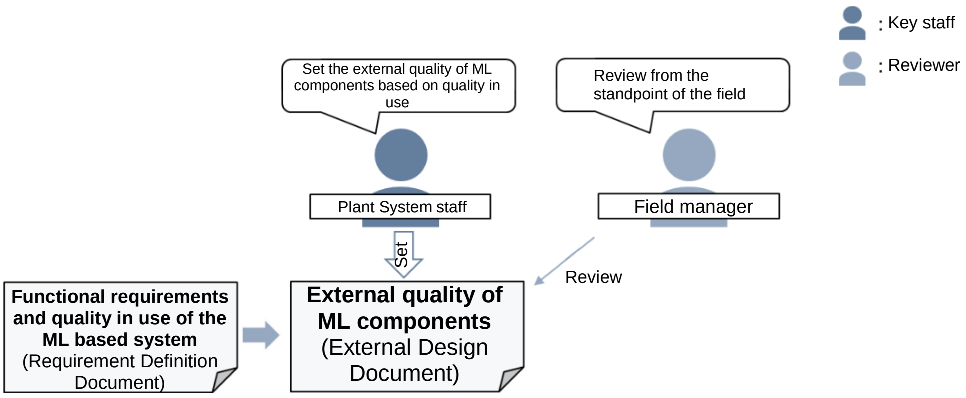


Figure 4-4 Conceptual image of implementation of Step 3 “Setting the external quality”

[91]At the stage of this step "Setting External Quality", it is not necessary to set numerical targets specific to ML (Example: Accuracy greater than or equal to ○%). In addition, regarding the external quality of "risk avoidance," it may be possible to define numerical targets required of ML components (e.g. the incidence of misjudgments leading to danger) by checking safety-related systems and external safety mechanisms in the process of Step 5 Setting the level of external quality. At the stage of building an ML component (Step 7 "Design and development of an ML component" in Chapter 4 of the Guidelines), concrete numerical targets specific to machine learning (e.g. accuracy, F-measure) can be set according to the results of PoC, data acquisition status, learning status, etc.

4) Organizing the ML based system and safety-related systems pertaining to the level of external quality

- Implementations

Plant System staff

Organize the components of the system (ML components, non-ML components (external safety mechanism, etc.)) necessary to achieve quality in use of the ML based system and to set the level of external quality of ML components, as well as the existence and functional requirements of safety-related systems independent of the system

Field manager

When there are non-ML components (external safety mechanism, etc.) of the ML based system or safety-related systems independent of the ML based system, review the components and safety management systems from the standpoint of the field manager

Environment & Safety staff

When there are non-ML components (external safety mechanism, etc.) of the ML based system or safety-related systems independent of the ML based system, review the components and safety management systems from the standpoint of safety management

- Relevant section of the Guidelines

Plant System staff shall confirm each component of the ML based system (ML components and non-ML components (external safety mechanism, etc.)), and the method for setting the level of external quality of ML components based on the relationship to safety-related systems independent of the system in "2.2.3 Setting the level of external quality," and organize the non-ML components of the system, as well as the existence and functional requirements of safety-related systems that are independent of the system[92]. The Field Manager and Environment & Safety staff will refer to the same section as needed.

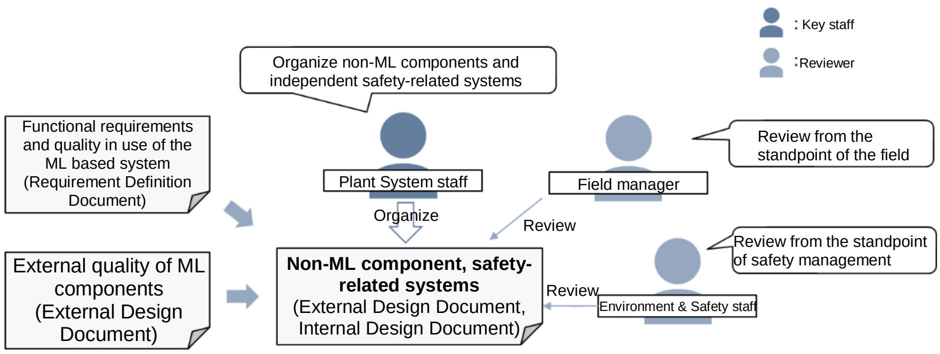


Figure 4-5 Conceptual image of implementation of Step 4 “Organizing ML based system and safety-related systems pertaining to the level of external quality”

[92] If it is confirmed that the safety-related systems independent of the ML based system are sufficiently safe based on the existing system development process such as the functional safety standards, IEC61508 (JIS C 0508) and IEC61511 (JIC C 0511), it is judged that there is no need to set the "risk avoidance" axis item in the ML components of the quality in use and external quality set in steps 2 and 3. The example is shown when the quality in use and external quality of the "risk avoidance" axis are not set in the use case of "3.3.3 Equipment deterioration diagnosis."

5) Setting the level of external quality[93]

- Implementations

<u>Plant System staff</u>
Set the level of external quality of ML components

<u>Field manager</u>
Review the level of external quality from the standpoint of the field manager

- Relevant section of the Guidelines

Plant System staff will understand how to set the external quality level in "2.2.3 Setting the level of external quality.” Based on the external quality of ML components and non-ML components of the ML based system (external safety mechanism, etc.), the external quality level is set according to the method described above. The risk avoidance is set by combining SIL assessment of the overall ML based system and the use of AISL Table as needed. Set the "Performance" to an appropriate level based on the criteria. The Field Manager will refer to the same section as needed.

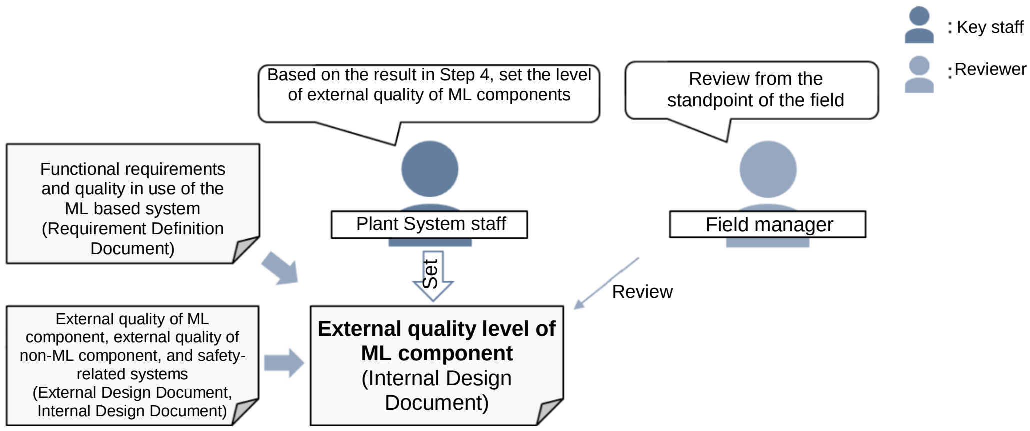


Figure 4-6 Conceptual image of implementation of Step 5 “Setting the level of external quality”

[93] Regarding numerical targets required for ML componets, in the process of this step "Setting the level of external quality", it may be possible to determine the rate of errors leading to danger by confirming safety related systems and external safety mechanisms. Finally, at the stage of building an ML component (Step 7 "Design and development of an ML component" in Chapter 4 of the Guidelines), concrete numerical targets specific to machine learning (e.g. accuracy, F-measure) can be set based on the results of PoC, data acquisition status, learning status, etc.

(4) Implementation

In implementing the ML based system, ML components and non-ML components are developed by realizing the internal quality. This is an important process to consider and implement specific measures in order to ensure the reliability of ML components.

Table 4-10 Implementations in the “implementation” phase

Step	Activity Item in Quality Assurance	Implementer	Relevant Section of the Guidelines
6	Setting the level of internal quality	ML Design/Development staff: Set the level of internal quality of ML components based on the level of external quality	2.1.4: Understand an overview of the eight axes of internal quality 2.2.4: Check the level setting of internal quality Appendix (Checklist): Check the perspectives in the field of plant safety for internal quality (Refer to 3.3 and Appendix (Checklist): Use case-specific perspectives in internal quality)
7	Design and development of ML components	ML Design & Development staff: Develop according to the design of ML components, internal quality requirements and perspectives ML Quality Assurance staff: In the development of ML components, check whether the requirements and perspectives of internal quality are satisfied from the standpoint of quality assurance (Field staff, Plant System staff: Review from the standpoint of data management or field users, and provide data)	2.2.5, Appendix (Checklist): Confirm the execution of internal quality "requirements" and "perspectives" and the specific items to be executed
8	Development of non-ML components of the ML based system (external safety mechanism, etc.)	Plant System staff: Develop the non-ML components required for external quality and quality in use (external safety mechanism, etc.)	-

6) Setting the level of internal quality

- Implementations

<u>ML Design & Development staff</u>

Set the level of internal quality of ML components according to the level of external quality, and check the requirements corresponding to the set level, as well as the perspectives in the field of safety and the use case-specific perspectives

- Relevant section of the Guidelines

ML Design & Development staff shall understand the contents of the eight axes of internal quality in "2.1.4 Internal quality" and how to set the level of internal quality in "2.2.4 Confirming the level of internal quality.” Set the level of internal quality of ML components according to the external quality of ML components and the level of external quality. In "Appendix (Checklist)," check the "requirements" and "perspectives in the field of plant safety" according to the level setting. If necessary, check the use case “Use case-specific perspectives” similar to the ML based system to be built this time in "3.3 Specific application of reliability assessment based on use cases." Plant System staff will refer to the same section as needed.

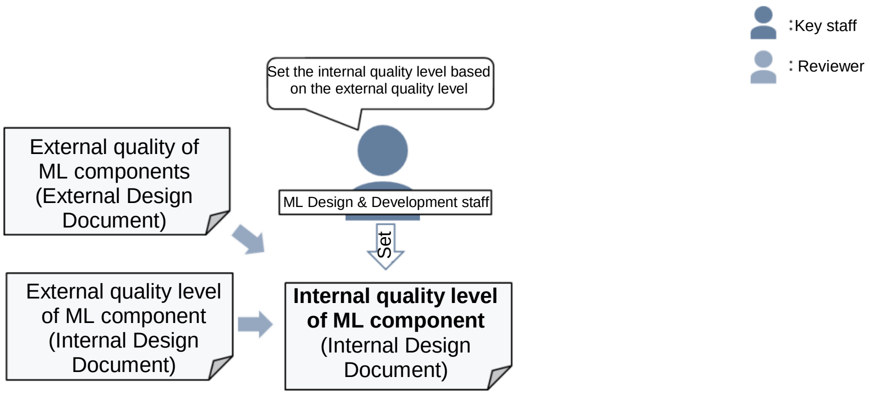


Figure 4-7 Conceptual image of implementation of Step 6 “Setting the level of internal quality”

7) Design and development of ML components[94]

- Implementations

ML Design & Development staff

Design the ML components (identification of concrete specifications and models of ML components); development based on internal quality requirements and perspectives

ML Quality Assurance staff

In the development of ML components, check whether the internal quality requirements and perspectives are satisfied from the standpoint of quality assurance

Field/Plant System staff

Review from the standpoint of data management, field and users in the development of ML components, provide data, and cooperate in labeling

- Relevant section of the Guidelines

ML Design & Development staff shall check the position of "requirements" and "perspectives" and how to execute them in "2.2.5 Checking and executing internal quality requirements," and design and develop ML components according to the "requirements" and "perspectives" in the "Appendix (checklist)" confirmed in Step 6. In the process, the ML Quality Assurance staff shall check whether the internal quality requirements and perspectives are satisfied based on the Guidelines and the corresponding section of the "Appendix (Checklist)," and keep the record of responses. Field and Plant System staff will refer to the same section as needed.

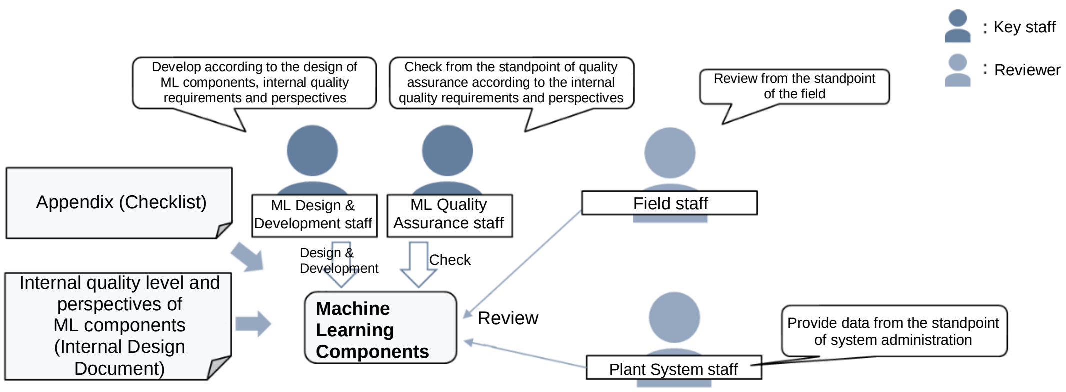


Figure 4-8 Conceptual image of implementation of Step 7 “Design and development of ML component”

[94] In this step "Design and development of ML component," concrete numerical targets specific to machine learning (e.g. accuracy, F-measure) can be set according to the results of PoC, data acquisition status, learning status, etc.

8) Development of non-ML components of the ML based system (external safety mechanism, etc.)

- Implementations

<u>Plant System staff</u>

Develop non-ML components of the ML based system (external safety mechanism, etc.) required to meet the quality in use and external quality requirements

- Relevant section of the Guidelines

This activity is necessary to achieve the quality in use by satisfying the requirements for the external quality level of ML components. However, since it is not directly related to the ML components, it is not included in the Guidelines[95].

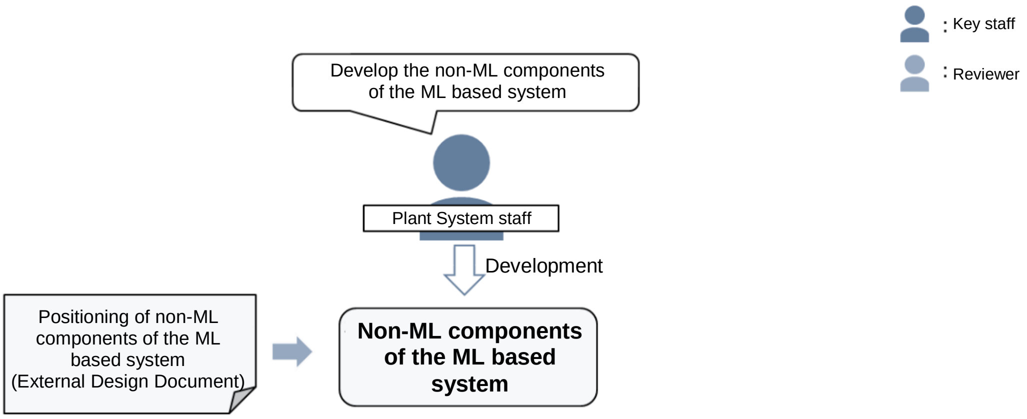


Figure 4-9 Conceptual image of implementation of Step 8 “Development of non-ML components of the ML based system (external safety mechanism, etc.)”

[95] Non-ML components (external safety mechanism, etc.) should be developed in accordance with existing system development processes, such as functional safety standards IEC61508 (JIS C 0508) and IEC61511 (JIC C 0511).

(5) Testing & acceptance

Test the ML based system and assess the results. In addition, an acceptance inspection is conducted based on the test results. Since the internal quality of ML components is confirmed in Step 9 "Design and development of ML components," this phase focuses on assessing the external quality of ML components in the ML based system. The acceptance inspection checks the requirements, including quality in use.

Table 4-11 Implementations in the “testing and acceptance” phase

Step	Activity Item in Quality Assurance	Implementer	Relevant Section of the Guidelines
9	Testing (assessment of external quality)	System Quality Assurance staff: Test the ML based system and assess the results. Assess whether the external quality of ML components meets the required level (Plant System staff: Based on the assessment of test results, review from the standpoint of having organized external safety mechanisms and safety-related systems and having set the external quality and its level Environment & Safety staff: Review the test results from the standpoint of safety management)	2.2.3: Check external quality assessment criteria
10	Acceptance inspection	System Quality Assurance staff: Assess overall ML based system based on the results of assessing external quality in the test, and if the ML based system satisfys the criteria based on the assessment results, etc., conduct an acceptance inspection (Project Planning staff: Review the overall ML based system from the standpoint of project originator Field staff: Review the quality in use assessment results from the standpoint of the field users in use from the standpoint of the field and users)	-

9) Testing (assessment of external quality)

- Implementations

<u>System Quality Assurance staff</u>
Test the ML based system and assess the results. Assess whether the external quality of ML components meets the required level

<u>Plant System staff</u>
Based on the assessment of test results, review from the standpoint of having organized external safety mechanisms and safety-related systems and having set the external quality and its level

<u>Environment & Safety staff</u>
Review the assessment on test results from the standpoint of safety management

- Relevant section of the Guidelines

System Quality Assurance staff should understand the assessment criteria and methods for external quality in "2.2.3 Setting the level of external quality." Based on the level and setting of external quality of ML components, assess whether the external quality of ML components in the test results meets the required level. Plant System staff and Environment & Safety staff will refer to the same section as needed.

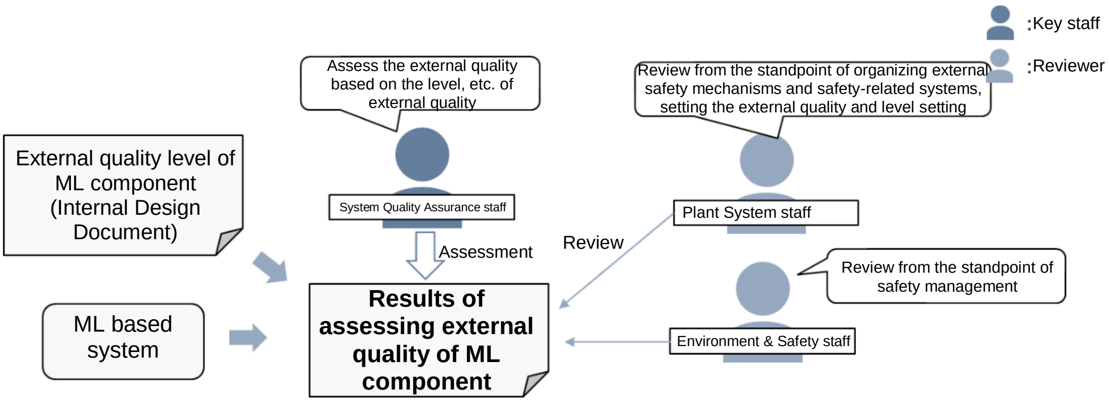


Figure 4-10 Conceptual image of implementation of Step 9 “Testing (assessment of external quality)”

10) Acceptance inspection

- Implementations

System Quality Assurance staff

Assess the overall ML based system based on the results of assessing external quality in the test, and if the ML based systemML based system satisfys the criteria based on the overall results, conduct an acceptance inspection

Project Planning staff

Review the overall ML based system from the standpoint of project originator

Field staff

Review the quality in use assessment results from the standpoint of the field users

- Relevant section of the Guidelines

Since the acceptance inspection is based on the standards of each company, there is no directly applicable part in the Guidelines. But it is necessary to appropriately follow the level of external quality, the level of internal quality set accordingly, the internal quality requirements and records of responses, and the results of assessing external quality in testing.

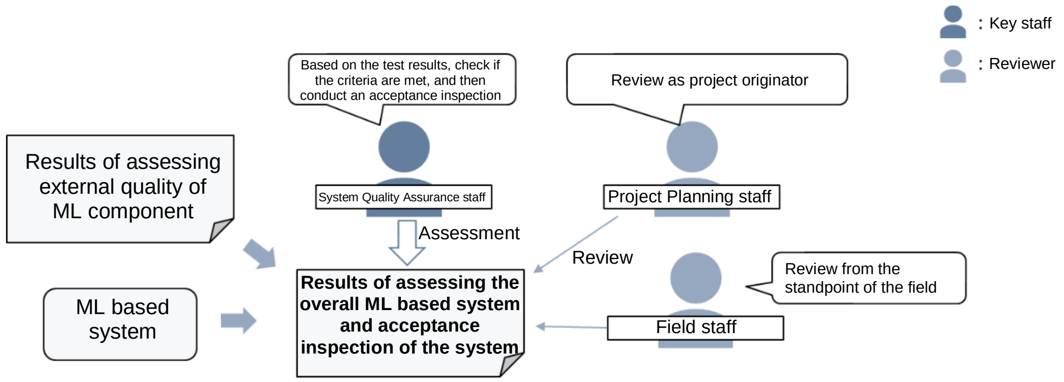


Figure 4-11 Conceptual image of implementation of Step 10 “Acceptance inspection”

(6) Operation

After the start of the ML based system operation, a system update is performed as needed. In the case of both updating after each quality check and updating in real time, it is necessary to monitor the output (quality in use/external quality) of ML components and the overall ML based system. In addition, a system update will be required based on the results, and in many cases, it is necessary to prepare the criteria and other mechanisms for such decisions before the start of operations (at the time of "(3) Design" phase)[96].
Based on the above, the activity items in the operation phase describes the quality assurance activities which are conducted based on the predetermined frequency and implementation standards.

Table 4-12 Implementations in the "operation" phase

Step	Activity Item in Quality Assurance	Implementer	Relevant Section of the Guidelines
11	Confirming the quality in use	System Quality Assurance staff: Confirm the quality in use of the ML based system in operation (Field staff: Review the results of confirmation of quality in use from the standpoint of the field and users)	2.1.1./2.1.3/2.2.1: Confirm the overview of quality in use
12	Confirming the external quality	System Quality Assurance staff: Confirm the external quality based on the results of confirmamton of quality in use of the ML based system in operation (Plant System staff: Review the results of confirmation of external quality from the standpoint of having organized external safety mechanisms and safety-related systems and having set the external quality and its level)	2.2.3: Check external quality assessment criteria
13	Internal Quality Confirming the [Maintainability of qualities in use]	ML Quality Assurance staff: Check the status of compliance with internal quality requirements and perspectives based on the results of confirmation of quality in use and external quality of the ML based system in operation	2.2.4: Check the internal quality assessment criteria 2.2.5, Appendix (Checklist): Check the internal quality requirements and perspectives

96 National Institute of Advanced Industrial Science and Technology (AIST) (2020), "1st edition of Guidelines on Quality Control for Machine Learning"

14	System update	ML Design & Development staff: Update ML components according to the results of the external quality and internal quality checks Plant System staff: Update the non-ML components (external safety mechanism, etc.), and provide data for updating ML components from the standpoint of system management (Field staff: Review from the standpoint of the field users in updating ML components)	2.2.4: Check the internal quality assessment criteria 2.2.5, Appendix (Checklist): Refer to the internal quality requirement and perspectives

11) Confirming the quality in use

- Implementations

System Quality Assurance staff

Check the achievement status of quality in use of the ML based system in operation at the predetermined timing

Field staff

Review the results of confirmation of quality in use from the standpoint of the field users

- Relevant section of the Guidelines

System Quality Assurance staff need to understand the position of quality in use in "2.1.1 Quality in use" and "2.1.3 Quality in use and external quality axes," and how to set quality in use in "2.2.1 Setting quality in use." Based on the functional requirements and quality in use of the system at the time of development, confirm that the quality in use of the system in operation meets the initial purpose. Field staff will refer to the same section as needed.

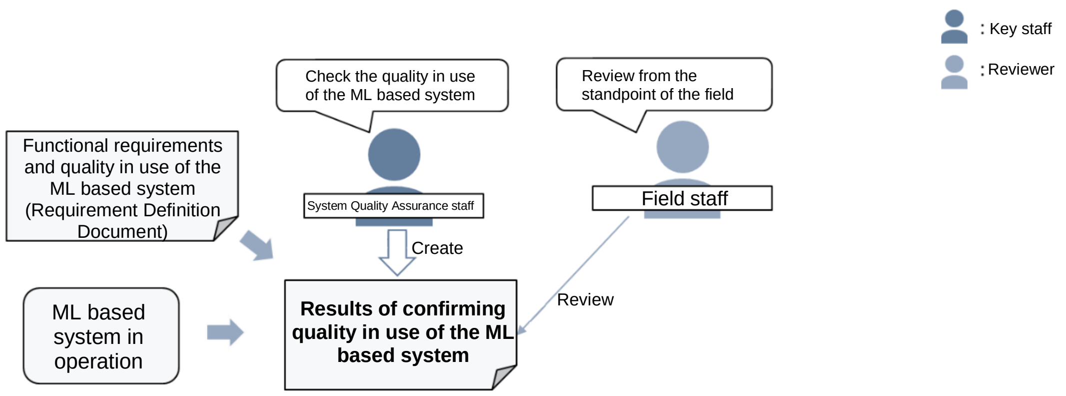


Figure 4-12 Conceptual image of implementation of Step 11 “Confirming the quality in use”

12) Confirming the external quality

- Implementations

System Quality Assurance staff

Confirm the achievement status of external quality of ML components based on the results of confirmation of quality in use of the ML based system in operation

Plant System staff

Review the results of confirmation of external quality from the standpoint of having organized external safety mechanisms and independent safety-related systems and having set the external quality and its level

- Relevant section of the Guidelines

System Quality Assurance staff should understand the assessment criteria and methods for external quality in "2.2.3 Setting the level of external quality.” Based on the functional requirements and quality in use of the ML based system and external quality and of its level of ML components, confirm that the external quality of ML components in operation meets the initial requirements. Plant System staff will refer to the same section as needed.

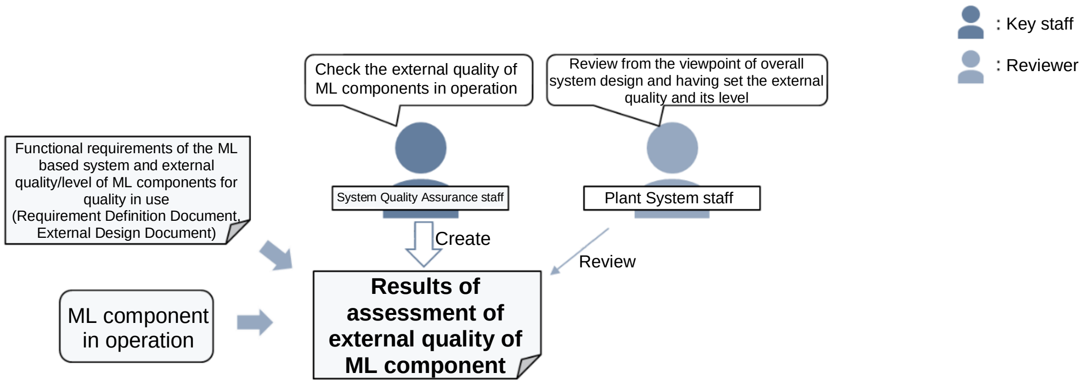


Figure 4-13 Conceptual image of implementation of Step 12 “Confirming the external quality”

13) Confirming the internal quality (Maintainability of qualities in use)

- Implementations

<u>ML Quality Assurance staff</u>

Check the status of compliance with internal quality requirements for "Maintainability of qualities in use"[97] based on the results of confirmation of quality in use and external quality of the system in operation

- Relevant section of the Guidelines

Based on "2.2.4 Confirming the level of internal quality," "2.2.5 Checking and executing internal quality requirements," and "Appendix (Checklist)," the ML Quality Assurance staff shall confirm whether the "requirements" and "perspectives" of internal quality are met, especially with regard to "Maintainability of qualities in use."

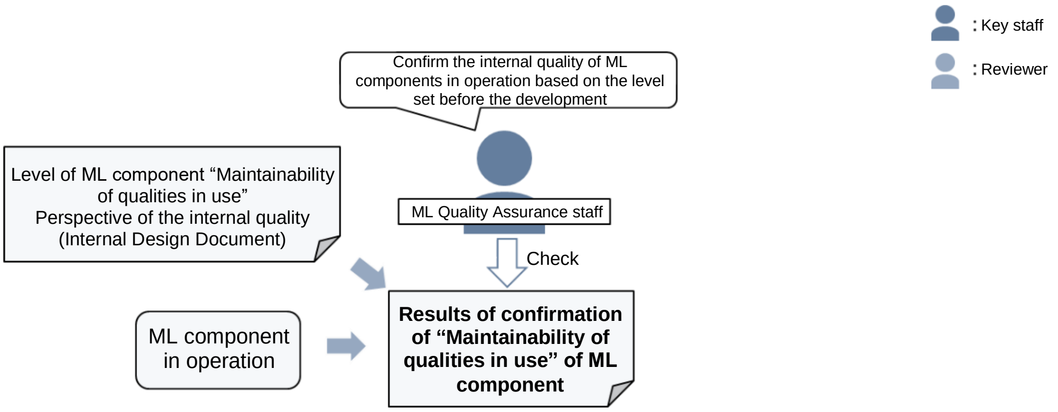


Figure 4-14 Conceptual image of implementation of Step 13 "Confirming the internal quality"

[97] Through the fulfillment of internal quality of "Maintainability of qualities in use," check that the achievement status of other internal qualities has not declined.

14) System update[98]

- Implementations

<u>ML Design & Development staff</u>
Update ML components according to the results of the external quality and internal quality checks

<u>Plant System staff</u>
Update the non-ML components (external safety mechanism, etc.) according to the results of confirmation of quality in use and external quality, and provide data for designing and development of ML components from the standpoint of system management

<u>Field staff</u>
Provide opinions as needed from the standpoint of the field users in updating ML components

- Relevant section of the Guidelines

ML Design & Development staff shall add corrections to the internal quality according to the results of confirmation based on "2.2.4 Confirming the level of internal quality," "2.2.5 Checking and executing internal quality requirements," and "Appendix (Checklist)," and update the ML components. Field and Plant System staff will refer to the same section as needed. There are no relevant sections in the Guidelines for updates of non-ML components.

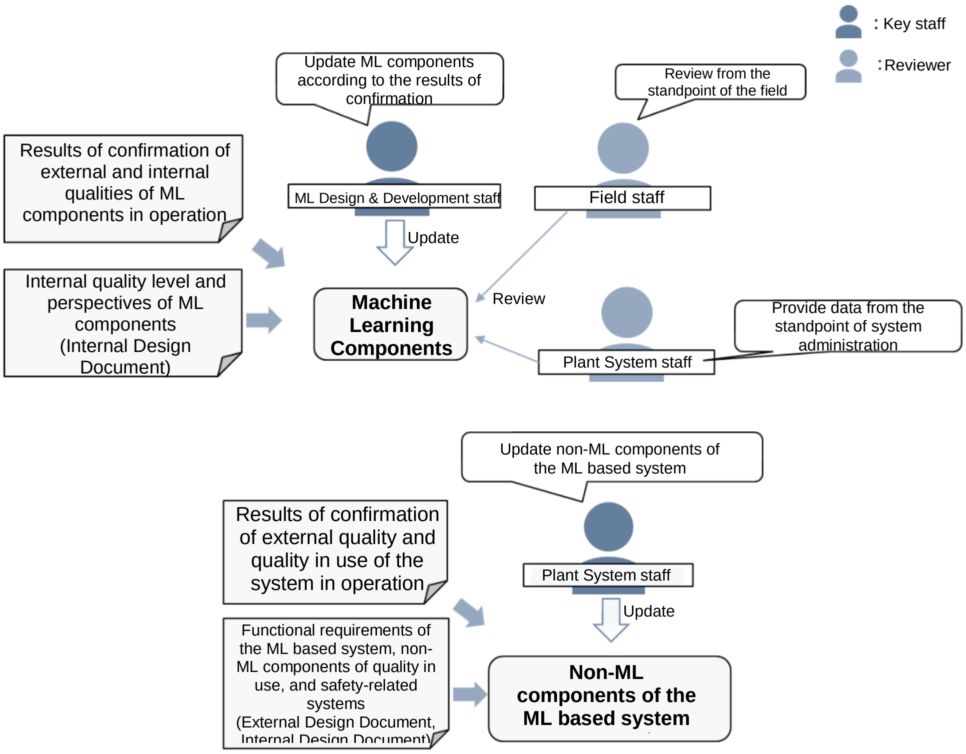


Figure 4-15 Conceptual image of implementation of Step 14 “System update”

[98] External quality of the updated system is confirmed in Step 9 (Testing) by returning to the testing and acceptance phase.

“Guidelines on Assessment of AI Reliability in the Field of Plant Safety”: FAQ

#	Classification	Question	Answer
1	Chapter 1	How do the Guidelines relate to the various laws and regulations currently applicable to plant operations and the required legal procedures?	The Guidelines are not intended to relax or interpret the provisions of laws and regulations, and it is necessary to comply with legal obligations when using ML components for statutory inspections. (See 1.3 Scope of Application) In 2020, a notification based on the High Pressure Gas Safety Act was revised to clarify that AI can be used for completion inspections, safety inspections, daily inspections, etc. The Notification mandates that safety be taken into consideration, by referring to materials such as the Guidelines. As for the conditions under which AI can be used in statutory inspections and checks, consulting the notification, ministerial ordinances, etc. is necessary.
2	Chapter 1	Are responses to cyber attacks and other external attacks included in the Guidelines?	We recognize that dealing with external attacks, including cyber security-related, is a priority issue for plant systems. However, it is notcovered specifically by the Guidelines, as the issue is not specific to machine learning, and is considered as a matter requiring separate examination. (See 1.3 Scope of Application)

#	Classification	Question	Answer
3	Chapter 2	How can one declare that "AI has been developed, operated, etc. by an appropriate method to ensure safety and avoid losses" by using the Guidelines?	The following three points must be met to demonstrate the reliability of AI, i.e. that the expected quality (safe and incurs less loss) will be achieved. 1. Performance required of AI is appropriately limited by non-AI components, such as safety-related systems independent of ML compoments, and operations, such as human involvement 2. Processes for training, testing, implementing, and operating AI are appropriately designed 3. The AI is sufficiently tested The Guidelines can be used to meet the aforementioned requirements by the method shown below. Point 1can be accounted for by setting external quality items and AISL/AIPL according to the procedure Point 2 can be accounted for by satisfying the internal quality requirements. By following the Guidelines and meeting the requirements 1 and 2, and if the desired performance is obtained through sufficient testing (thereby satisfying point 3), all three requirements can be satisfied.
4	Chapter 2	Is there a guideline for the "performance levels" when setting AIPL?	Levels of performance are not set indiscriminately in the Guidelines, but are set by mutual agreement between the user and vendor companies.
5	Chapter 2	Should the items described in "quality in use" and "external quality" include functions to be implemented in the future?	The quality in use and external quality items should not be the goals desired to be achieved in the future, but rather the goals to be achieved in the current development phase. When determining AISL for the item, it is useful to consider in advance the potential incrementation of AISL regarding future goals (e.g. a system may currently instruct humans to check the output, but may have human interactions limited and automate such actions in the future). (Making note of such points in the remarks column of the template for future consideration is advised.)

#	Classification	Question	Answer
6	Chapter 2	When there is no room for error in AI, should "zero errors" be set as the external quality?	With machine learning, it is theoretically difficult to achieve an accuracy of 100%. Therefore, it is important to ensure safety in combination with existing systems, and not to entrust the ML components with extensive safety functions. Consult the column titled "Is '100% accuracy' necessary for AI?" at the end of Chapter 2 of the Guidelines, which explains related topics. In addition, external quality should not be a conceptual goal such as "zero errors," but a realistic goal that is required of the ML component. The Guidelines provide expressions such as "minimize the rate of false negatives" as examples in use cases.
7	Chapter 2	How and when should quantitative performance targets be set for AI?	At the stage of building an ML component (Step 7 "Design and development of an ML component" in Chapter 4 of the Guidelines), concrete numerical targets specific to machine learning (e.g. accuracy, F-measure) can be set according to the results of PoC, data acquisition status, training status, etc. In addition, regarding the external quality of "risk avoidance," it may be possible to define numerical targets required of ML components (e.g. the incidence rate of potentially dangerous misjudgment) by checking safety-related systems and external safety mechanisms in the process of Step 5 "Setting the level of external quality". At the stage of Step 3 "Setting the external quality," a numerical target (e.g. accuracy of more than x%) for the ML component is not necessary.
8	Chapter 2	Apart from the accuracy of AI, it is also important that the ML based system is user-friendly (e.g. for workers). How should such viewpoint be handled in the Guidelines?	We believe that user-friendliness is included in the quality in use However, note that whether a system is subject to the Guidelines is determined based on whether it involves machine learning. If related to the quality of an ML component output (output content, timing, etc.), user-friendliness can be treated as "performance." On the other hand, the user interface of a system is outside the scope of the Guidelines.

#	Classification	Question	Answer
9	Chapter 2	We, a vendor, are proposing AI to clients. It is difficult for vendors to assess the impact of misjudgments of AI, which is necessaryfor AISL judgment. Is the involvement of the plant owner a requirement?	The reliability assessment in the Guidelines is based on the assumption that a system will be implemented in plants. Therefore, even if a vendor leads the project, it is essential that the plant owner is involved and evaluate the impact of AI.
10	Chapter 4	Is the involvement of “staff" mentioned in Chapter 4 mandatory? There are some aspects that do not fit with our AI development system.	Types of companies and departments that participate in ML based system development projects, as well as their respective roles, vary by project. There is no problem in making records according to actual application. Please also refer to "Practical Examples" of the Guidelines, where the staff members are listed based on actual reliability assessment cases conducted in real projects.
11	Chapter 4	Regarding the “staff” described in Chapter 4, what is the difference between Key Staff and Reviewer? From the examples given in the Guidelines, will there be any problem if the roles of Key Staff and Reviewer are switched?	Suppose a situation where the functional requirements and quality in use of a ML based system is set by the project planning staff (leading the project), and checks are done by a field staff (who will eventually use the system as part of their work). In this case, the project planning staff will be the Key Staff, and the field staff will be the Reviewer. The Reviewer need not read and understand the contents of the Guidelines, but should be involved in the reliability assessment in response to a request from the Key Staff. (Refer to Section 4.1 “Key players in utilizing the Guidelines”) Since roles may vary depending on the project, consider assignments described in Chapter 4 as examples for reference only. It is important to assign the roles of Key Staff and Reviewer appropriately to clarify who is involved in setting the quality. Refrain from simply setting the user company as Key Staff, and the vendor company as the Reviewer.
12	Others	How should the Guidelines be applied when multiple AIs are linked?	In the Guidelines, each AI is assessed individually. There are no special considerations for the output of one AI becoming the input of another.

#	Classification	Question	Answer
13	Others	Do the Guidelines apply to the introduction of an AI where the training is already complete?	A trained AI is also to be assessed by the methods described in the Guidelines, as conventional methods cannot assess the reliability of output from an AI, regardless of whether the training is already complete.
14	Others	If there are multiple plants to which an AI is applied, is it necessary to sum up the impacts of AI misjudgments for all the application sites and use the whole impact in AISL judgment?	In the Guidelines, assessment is based on the principle of "one implementing site at a time," so in the case of multiple sites implementation, reliability assessment should be conducted separately for each plant. If the impact of misjudgment could spread to multiple plants, it shall be treated as "the incidence of impacts at a single implementation site is more frequent.” However, the frequency of occurrence is not taken into account in the AISL Table (the AISL Table assumes a uniformly high frequency of occurrence). This is a matter to be considered when conducting a detailed assessment.
15	Others	Does having redundancy by simultaneously using multiple AIs lead to higher reliability?	At this point, we believe that it is difficult to secure redundancy by having multiple AIs. For example, an AI that aims to improve accuracy through via ensemble learning will be assessed as a single AI in the Guidelines. In the case of redundancy between an AI and a non-AI systems, the latter can be handled as "external safety mechanism" of AI, as long as the reliability of the non-AI system is confirmed.
16	Others	In actual AI development, there are many cases where the purpose of the system is not clear at the beginning. Are the Guidelines also useful in such cases?	The Guidelines are intended for consideration after the challenges faced by the user are fixed. Searching for such issues is a step before the reliability assessment.. Consulting the “Collection of Advanced AI Case Examples at the Plants” is advised, as it describes the issue of "difficulty in setting the objective for AI projects" and the solutions to overcome this difficulty, though the collection is available in Japanese only.

#	Classification	Question	Answer
17	Others	External quality is set for each individual AI development case, while internal quality requirements are set in advance and remain unchanged. Is external quality set only for the purpose of setting the level and not related to internal quality?	External quality and internal quality are closely interrelated. Internal quality requirements are set in advance in the Guidelines, but the methods for achieving those requirements (specific measures to ensure internal quality, as to be described in the sub-template) vary depending on the external quality, as well as its prerequisites, viz. quality in use and functional requirements. In other words, not only do the levels of internal quality requirements change according to the level of external quality, but the method of achieving internal quality is also different in each individual case, even when internal quality levels are the same.
18	Others	How can I use the checklist of internal quality requirements to show that the AI I am developing has no problems?	Internal quality requirements can be interpreted as having ensured the reliability of AI on the designated level. This would be achieved by addressing all requirements of a given level (Lv 1 – 3), either by meeting the requirements or provide reasons as to why they are not applicable. Internal quality Lv1 corresponds to AISL0.1 and AIPL1, internal quality Lv2 corresponds to AISL0.2 and AIPL2, and internal quality Lv3 corresponds to AISL1. For example, by addressing all the requirements of internal quality Lv2, one can describe a system as satisfying AISL0.2 or AIPL2.

Appendix

"Perspectives in the Field of Plant Safety" Checklist for Internal Quality Assurance

“Perspectives in the Field of Plant Safety” Checklist for Internal Quality Assurance

[How to use the checklist and template]

0 (Before looking at the checklist) Check the required level of internal quality (Level 1–3) in developing ML components. (Section 2.2.4 of the Guidelines)

1 Confirm the requirements of the applicable required level for the internal quality axis to be considered in accordance with the "Situation of use of the checklist" below.

2 Confirm the "Perspectives in the field of plant safety" directly related to the requirements of the applicable required level. Check the applicable required level (Lv1–Lv3) in the "Required level of internal quality" column.

When an ML component to be developed is similar to use cases, also confirm “Use case-specific perspectives.”

3 Confirm common "Perspectives in the field of plant safety" that apply regardless of requirements. In the "Required level of internal quality" column, it is stated as "Common."

When an ML component to be developed is similar to use cases, also confirm “Use case-specific perspectives.”

4 At the time of development and implementation, record the measures taken and the date when such measures were taken in the "Record of measures taken (for development and implementation)."

5 After the start of operation, at the time of system update, record the measures taken for quality confirmation and update in the "Record of measures taken for operation" of "Maintainability of qualities in use."

***For entry examples, "Practical Examples," in which this record template is based on actual cases of AI development and operation for each use case, are also available. Consulting these practical examples is advised.**

[Situation of use of the checklist and template]

data quality desig	1 Consideration of "data quality design" (Sufficiency of requirement analysis and Coverage for distinguished problem cases)
	Example of consideration:
	・Consideration of attributes and environmental factors to be covered in data collection
	・Confirmation of requirements analysis and collected data with on-site engineers, etc.
Data check	2 Consideration of "Data quality" (Coverage of datasets and Uniformity of datasets)
	Example of consideration:
	・Confirmation that attributes and environmental factors to be covered have been collected
	・Consideration of the use of simulators when data is insufficient, etc.
Learning	3 Consideration of "Model quality" (Correctness of the trained model and Stability of the trained model)
	Example of consideration:
	・Consideration of allowable level of misjudgment
	・Consideration of test methods, etc.
Implementation ar	4 Consideration of "Quality of implementation and operation" (Reliability of underlying software systems and Maintainability of qualities in use)
	Example of consideration:
	・Consideration of software used for implementation
	・Frequency design of accuracy verification and re-learning, etc.

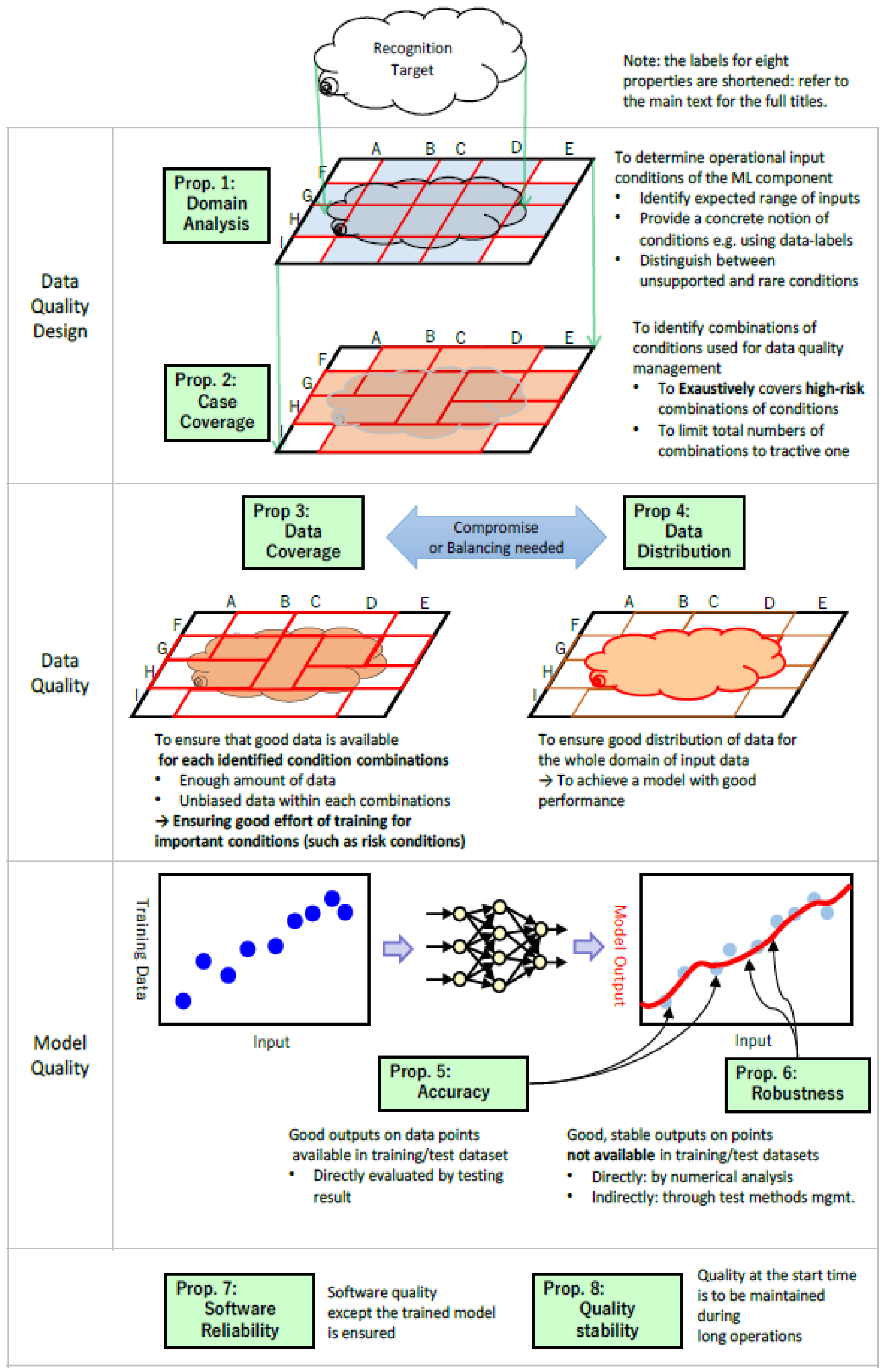


Figure: Internal Quality in “Guidelines on Quality Control for Machine Learning”

Source: National Institute of Advanced Industrial Science and Technology (AIST) “Machine Learning Quality Management Guideline 1st edition ”

“Perspectives in the Field of Plant Safety” Checklist for Internal Quality Assurance

Situation of use	Internal quality requirements in “Machine Learning Quality Management Guideline 1st edition”				Perspectives in the field of plant safety	Related “Use case-specific perspectives” (see Section 3.3 of the text)				
	Internal quality axis	Required level of internal quality	Requirement No.	Internal quality requirements		Prediction of pipe wall thickness	Pipeline diagnostic imaging	Equipment deterioration diagnosis	Detection and diagnosis of early signs of abnormality	Optimization of operation
data quality design	1 Sufficiency of requirement analysis	Lv1	1	Examine and record the major cause of possible deterioration of quality	・Did you consider the following as "Major risks of quality deterioration" and "Causes thereof" in the plant field? - Environmental changes: Season, weather, day and night, temperature, location, etc. - Changes in product characteristics: Types, components, etc. - Changes in the state of plant: Start-up, normal operation, etc.	—	—	—	—	—
data quality design	1 Sufficiency of requirement analysis	Lv1	2	Based on the examination results, design data and reflect it in necessary attributes.	・Did you design data based on the results of assessment of "Causes of possible deterioration of quality" in the plant field?	—	—	—	—	—
data quality design	1 Sufficiency of requirement analysis	Lv2	3	Analyze risks of deterioration of quality in use in overall system and their impact with a certain level of engineering coverage and record the results in documents.	・As for analysis with a certain engineering coverage, did you use existing information (if any) related to SIL assessment of overall safety-related systems and engineering risk analysis such as FTA, STAMP/STPA? ・When existing engineering analysis is not available, did you perform a new analysis with a certain degree of coverage?	—	—	—	—	—
data quality design	1 Sufficiency of requirement analysis	Lv2	4	Analyze if any measure is required for each of those risks, and analyze attributes related to the risk which are contained in an input to machine learning components.	—	—	—	—	—	—
data quality design	1 Sufficiency of requirement analysis	Lv2	5	Analyze and record the application-specific characteristics of environments which will generate machine learning input, with regards to the difficulty for machine learning and other aspects.	—	—	—	—	—	—
data quality design	1 Sufficiency of requirement analysis	Lv2	6	Examine sets of attributes and attribute values, based on the results of those analysis and record the background of such decisions.	—	—	—	—	—	—
data quality design	1 Sufficiency of requirement analysis	Lv3	7	The following activities are carried out in addition to those listed in Lv 2.	—	—	—	—	—	—
data quality design	1 Sufficiency of requirement analysis	Lv3	8	Investigate documents on own past examination results and those of others with regard to elements to be extracted as characteristics of system environment and record the background of examinations leading to the extraction of necessary subsets.	・Did you consider the following as elements to be considered as characteristics of the usage environment" in the plant field? - Environmental changes: Season, weather, day and night, temperature, location, etc. - Changes in product characteristics: Types, components, etc. - Changes in the state of plant: Start-up, normal operation, etc.	—	—	—	—	—
data quality design	1 Sufficiency of requirement analysis	Lv3	9	Investigate past examination results in line with application fields of systems with regard to deterioration risks of qualities in use of overall systems and record the examination results including the background of selection.	—	—	—	—	—	—
data quality design	1 Sufficiency of requirement analysis	Lv3	10	Moreover, extract deterioration risks of qualities in use of overall systems using engineering analysis such as Fault Tree Analysis and record their results.	・As “engineering analysis,” did you conduct SIL assessment of overall safety-related systems and an analysis of engineering risks such as FTA, STAMP/STPA? (mandatory for Lv3)	—	—	—	—	—
data quality design	1 Sufficiency of requirement analysis	Common	11	—	・Did you have plant engineers on site conduct a requirements analysis to analyze whether all usage conditions at the plant are covered?	—	—	—	—	—
data quality design	1 Sufficiency of requirement analysis	Common	12	—	・In the case of AI that uses a camera for recognition, did you narrow down the locations and conditions of the equipment to be recognized?	*This use case does not apply when recognition is not performed by a camera or other means.	・ When handling piping that is wrapped with insulation, be aware that the deterioration of insulation will be the target, not the deterioration of the piping itself.	*This use case does not apply when recognition is not performed by a camera or other means.	*This use case does not apply when recognition is not performed by a camera or other means.	*This use case does not apply when recognition is not performed by a camera or other means.

“Perspectives in the Field of Plant Safety” Checklist for Internal Quality Assurance

Situation of use	Internal quality requirements in “Machine Learning Quality Management Guideline 1st edition”				Perspectives in the field of plant safety	Related “Use case-specific perspectives” (see Section 3.3 of the text)				
	Internal quality axis	Required level of internal quality	Requirement No.	Internal quality requirements		Prediction of pipe wall thickness	Pipeline diagnostic imaging	Equipment deterioration diagnosis	Detection and diagnosis of early signs of abnormality	Optimization of operation
data quality design	1 Sufficiency of requirement analysis	Common	13	—	・In the case of AI that detects and predicts changes in equipment conditions, did you narrow down the locations and conditions to be detected and predicted?	・ As a type of corrosion affects the development of "Coverage for distinguished problem cases" and "Coverage of datasets," narrow down the scope to the specific type of corrosion.	—	・Determine the range of component values of products to be considered that vary with processing conditions. This includes not only the case where the product to be processed is different, but also the case when the fluid* and the process changes. *Changes in the distribution of mixed flow/multi-phase flows, etc.	・ Since the evaluation of "Coverage for distinguished problem cases" and "Coverage of datasets" is affected by types and location of targeted abnormality, specify the requirements including t y pes and location.	—
data quality design	1 Sufficiency of requirement analysis	Common	14	—	・When the requirements include an explanation of engineering cause and effect, did you check to see if it is mandatory for the ML based system?	—	—	—	・ Even if a causal relationship in an engineering sense between the detection of an abnormality and the related variables is not accounted for, it is acceptable to use the suggested relationship for inferrence.	—
data quality design	1 Sufficiency of requirement analysis	Common	15	—	・Can you secure the amount of data not only for learning but also for cross-validation and a generalization performance check?	—	—	—	—	—
data quality design	1 Sufficiency of requirement analysis	Common	16	—	・Do user companies provide data that will help you resolve issues? Or, is it possible to generate and acquire such data?	—	—	—	—	—
data quality design	1 Sufficiency of requirement analysis	Common	17	—	・Is the number of explanatory variables and causality in the learning data set too complex or too simple to model an issue? Do you also consider multicollinearity?	—	—	—	—	—

“Perspectives in the Field of Plant Safety” Checklist for Internal Quality Assurance

Situation of use	Internal quality requirements in “Machine Learning Quality Management Guideline 1st edition”				Perspectives in the field of plant safety	Related “Use case-specific perspectives” (see Section 3.3 of the text)				
	Internal quality axis	Required level of internal quality	Requirement No.	Internal quality requirements		Prediction of pipe wall thickness	Pipeline diagnostic imaging	Equipment deterioration diagnosis	Detection and diagnosis of early signs of abnormality	Optimization of operation
data quality design	1 Sufficiency of requirement analysis	Common	18	—	・Regarding the input data in operation, has the necessity of a mechanism to detect and eliminate inappropriate data that may lead to abnormal behavior been considered? (Example: Input data in operation obtained from a population different from the learning data in learning and outliers of input data.)	—	—	—	—	—
data quality design	1 Sufficiency of requirement analysis	Common	19	—	・Do you have a process or system in place that allows you to incorporate the experience of existing AI applications as a technology into the next development?	—	—	—	—	—
data quality design	2 Coverage for distinguished problem cases	Lv1	20	Set cases for each of attributes corresponding to major risk factors.	—	—	—	—	—	—
data quality design	2 Coverage for distinguished problem cases	Lv1	21	Moreover, set cases corresponding to combinations of composite risk factors.	—	—	—	—	—	—
data quality design	2 Coverage for distinguished problem cases	Lv1	22	Furthermore, extract attributes of differences in particularly-important environmental factors and prepare cases corresponding to combinations with serious risk factors.	・Did you extract "Environmental factors" in the plant field? - External environment: Weather, temperature, location, etc. - Production process: Production load, operating procedures, etc.	・ "Environmental factors" in this context refer to climate, salinity (regional characteristics such as distance from the sea and wind direction), and others.	・ "Environmental factors" in this context refer to sunlight, weather, seasons, time of day, and others. ・ In order to cope with image blurring, it is possible to absorb it by using a model, but ascertain the possibility of increasing complexity and uncertainty of the system.	・ "Environmental factors" in this context refer to location, operating environment, temperature and humidity, operating method, raw materials, utilities, etc.	・ "Environmental factors" in this context refer to factors which affect the detection of abnormality (e.g. production load, production lot).	・ "Environmental factors" in this context refer to operating procedures, raw materials, and others.
data quality design	2 Coverage for distinguished problem cases	Lv2	23	Satisfy all requirements listed in Lv 1.	—	—	—	—	—	—
data quality design	2 Coverage for distinguished problem cases	Lv2	24	Particularly-important risk factors should satisfy, in principle, the standards for pair-wise coverage. To be more specific, a case of combining “an attribute value of combination of those factors” and “individual attribute values included in all attributes other than those to which the attribute value belongs” should be included.	—	—	—	—	—	—
data quality design	2 Coverage for distinguished problem cases	Lv3	25	Based on engineering consideration, set standards for coverage of attributes and establish sets of combinations of attribute values that satisfied standards for coverage as cases.	—	—	—	—	—	—
data quality design	2 Coverage for distinguished problem cases	Lv3	26	The level of strictness of the standards for coverage (pair-wise coverage, triple wise coverage, etc.) should be set taking into account system usage and risk severity. Standards can be set individually for each risk where necessary.	—	—	—	—	—	—
data quality design	2 Coverage for distinguished problem cases	Common	27	—	・Did plant engineers on site check the case to see if the attributes related to the risk factors were extracted?	—	—	—	—	—
data quality design	2 Coverage for distinguished problem cases	Common	28	—	・In the case of AI that uses a camera for recognition, did you consider that the range of data and the ease of data acquisition will change depending on the location and material of the equipment to be recognized?	*This use case does not apply when recognition is not performed by a camera or other means.	・ As the color of the piping itself may be covered due to painting or anti-rust painting, ensure accuracy by considering these differences. ・ Keep in mind that there are cases where it is not possible to directly confirm the outer surface of the piping by images, such as when there is snow accumulation on the piping.	*This use case does not apply when recognition is not performed by a camera or other means.	*This use case does not apply when recognition is not performed by a camera or other means.	*This use case does not apply when recognition is not performed by a camera or other means.
data quality design	2 Coverage for distinguished problem cases	Common	29	—	・Have you developed rules so that the quality of data is kept constant?	—	・ Consider keeping the data quality at a certain level by establishing rules and points to consider for photographing.	—	—	—

“Perspectives in the Field of Plant Safety” Checklist for Internal Quality Assurance

Situation of use	Internal quality requirements in “Machine Learning Quality Management Guideline 1st edition” Internal quality axis	Required level of internal quality	Requirement No.	Internal quality requirements	Perspectives in the field of plant safety	Related “Use case-specific perspectives” (see Section 3.3 of the text) Prediction of pipe wall thickness	Pipeline diagnostic imaging	Equipment deterioration diagnosis	Detection and diagnosis of early signs of abnormality	Optimization of operation
data quality design	2 Coverage for distinguished problem cases	Common	30	—	・Did you consider the possibility of increasing system complexity and uncertainty if the model absorbs data quality fluctuations?	—	・ Blurred images may be included in training dataset so that the model can make decisions for such images as well. In such case, potential increase in system complexity and uncertainty should be put into consideration.	—	—	—
data quality design	2 Coverage for distinguished problem cases	Common	31	—	・In the case of AI that learns data on the characteristics (type, ingredients, etc.) of the products produced at the plant, has the possibility of collecting data covering the range of products been considered?	—	*This use case does not apply when data on product characteristics (type, contents, etc.) are not utilized.	・ Consider whether the training data can be collected for a range of component values for the target product.	—	—
data quality design	2 Coverage for distinguished problem cases	Common	32	—	・When using simulator data, did you check whether the simulator takes into account changes in environmental factors?	—	—	・When using simulation data, check whether the simulator takes into account changes in environmental factors (e.g. high humidity to low humidity). ・When obtaining data sets by simulation, fully verify the validity of the simulator.	・ When obtaining data sets by simulation, fully verify the validity of the simulator.	・ When obtaining data sets by simulation, fully verify the validity of the simulator.
data quality design	2 Coverage for distinguished problem cases	Common	33	—	・Did you consider the possibility that trends in data collection may change immediately after maintenance?	—	—	・The data immediately after a replacement of parts may be "No deterioration." Note that the "No deterioration" period depends on the specifications of the parts and materials, but varies depending on the usage environment (determine the period of "No deterioration" by referring to the frequency of past replacements, etc.). ・ When "running-in " is required immediately after maintenance of the equipment, ensure that no data are collected during that period.	—	—
Data check	3 Coverage of datasets	Lv1	34	Consider the source and method of acquiring test datasets to ensure that no bias is found in application situations.	・Did you extract "Application statuations" in the plant field?	・ "Application status" in this context refers to the targeted piping and the frequency of observations, the time axis of evaluation (e.g. whether to make a real-time projection), and others.	—	・ "Application status" in this context refers to the type and operation status (e.g. constant/temporary, load change) of the target equipment, and others.	・ "Application status" in this context refers to the severity of an abnormality to be detected and the situation of use of an ML based system (e.g. regular/temporary, daytime/night, steady/unsteady).	・ "Application status" in this context refers to the situation of operation to be applied (e.g. season, time of day, steady/unsteady, startup/shutdown in case of unsteady), equipment to be operated, and others.
Data check	3 Coverage of datasets	Lv1	35	Extract samples without bias from original data for each case to ensure that no bias is found.	・In the case of a framework that learns "normal" and classifies and predicts "abnormal," did you consider the difficulty of exhaustively covering abnormal data as test data?	—	—	—	・ In this case, it is not mandatory to cover the data of all casesof abnormality as training data. On the other hand, exhaustive sample extraction in normal domain is required.	—
Data check	3 Coverage of datasets	Lv1	36	Record activities carried out to prevent bias from entering.	—	—	—	—	—	—
Data check	3 Coverage of datasets	Lv1	37	Check that there are sufficient training data and test data for each analyzed case in the training phase, validation phase, and so on.	—	—	—	—	—	—
Data check	3 Coverage of datasets	Lv1	38	When sufficient training data cannot be acquired for any case, review and loose the coverage standards and record what should be checked individually by system integration tests in line with the original standards.	・When the amount of data in a specific range is not sufficient, have you considered the possibility that the classification and prediction accuracy of that range could become low?	・ When the amount of data in a specific range is not sufficient, recognize that the prediction accuracy of that range could become low.	・ When the amount of data in a specific range is not sufficient, keep in mind that the prediction accuracy of that range could become low.	・ Recognize that if sufficient operation data for a certain state cannot be obtained, the accuracy of detecting deterioration deviating from that state may be reduced.	・ Recognize that if sufficient normal data for a certain range cannot be obtained, the accuracy of detecting abnormality within that range may be reduced.	—
Data check	3 Coverage of datasets	Lv2	39	The following activities are carried out in addition to those listed in Lv 1.	—	—	—	—	—	—
Data check	3 Coverage of datasets	Lv2	40	Grasp an approximate probability of occurrence for each attribute value or each case.	—	—	—	—	—	—
Data check	3 Coverage of datasets	Lv2	41	Check if acquired data is not deviated from the distribution.	—	—	—	—	—	—
Data check	3 Coverage of datasets	Lv2	42	Positive check other than acquisition methods should be made regarding the coverage of the data included in each case.	—	—	—	—	—	—

“Perspectives in the Field of Plant Safety” Checklist for Internal Quality Assurance

Situation of use	Internal quality requirements in “Machine Learning Quality Management Guideline 1st edition”				Perspectives in the field of plant safety	Related “Use case-specific perspectives” (see Section 3.3 of the text)				
	Internal quality axis	Required level of internal quality	Requirement No.	Internal quality requirements		Prediction of pipe wall thickness	Pipeline diagnostic imaging	Equipment deterioration diagnosis	Detection and diagnosis of early signs of abnormality	Optimization of operation
Data check	**3 Coverage of datasets**	**Lv2**	43	For example, in each case, when there is any attribute not included in that case, extract the distribution related to attribute and check if there is no significant bias.	—	—	—	—	—	—
Data check	**3 Coverage of datasets**	**Lv3**	44	Acquire certain indicators for coverage of data included in each case in addition to those listed in Lv 2.	—	—	—	—	—	—
Data check	**3 Coverage of datasets**	**Lv3**	45	For example, check if there is no correlation between data other than attribute values included in combinations of cases using feature extraction or any other technique.	—	—	—	—	—	—
Data check	**3 Coverage of datasets**	**Lv3**	46	Or consider an expected distribution of attributes not included in each case, and analyze and record differences.	—	—	—	—	—	—
Data check	**3 Coverage of datasets**	**Common**	47	—	・Did plant engineers on site confirm whether the source of the data set was correct?	—	—	・ Personnel with expertise who can make appropriate judgments confirm whether the labels of detereoration are correct.	—	—
Data check	**3 Coverage of datasets**	**Common**	48	—	・Did you consider the handling of data in an unsteady state, such as when starting up a plant system?	—	—	—	—	—
Data check	**3 Coverage of datasets**	**Common**	49	—	・Did you consider the need to cover a wide range of data in an operating state because the state of a chemical plant changes constantly?	—	—	—	—	—
Data check	**3 Coverage of datasets**	**Common**	50	—	・When obtaining data by simulation, did you fully examine the validity of the simulator?	—	—	—	—	—
Data check	**3 Coverage of datasets**	**Common**	51	—	・Does the data set incorporate the effect of "disturbance" such as weather?	—	—	—	—	・ Pay attention to whether the assumed data set includes "disturbance" such as weather.
Data check	**3 Coverage of datasets**	**Common**	52	—	・Does the data set cover the data range of the assumed attributes?	・ Pay attention to whether the range of data of assumed attributes, such as contents, flow rate, material, flow velocity and pressure of piping, is covered.	・ Consider measures to deal with blurred input images, such as image blurring due to the surrounding environment (e.g. sunlight, time) or in drone photography. ・ Pay attention to whether the data range of each attribute of environmental factors is covered.	—	—	—
Data check	**3 Coverage of datasets**	**Common**	53	—	・When dealing with data under normal conditions, did personnel with expertise who can make appropriate judgments confirm that the data was actually under normal conditions?	—	—	—	・ Personnel with expertise who can make appropriate judgments confirm that the data under normal conditions is actually data under such conditions.	—
Data check	**3 Coverage of datasets**	**Common**	54	—	・When learning human operations, procedures, etc., isn't the case setting biased?	—	*This use case does not apply when data related to human operations and procedures are not utilized.	—	—	・ When learning the operations of experienced operators, make sure that the case setting is not biased.
Data check	**3 Coverage of datasets**	**Common**	55	—	・Did you check the basic statistics (e.g. percentage of missing values/outliers, mean/variance/covariance) for the data?	—	—	—	—	—
Data check	**3 Coverage of datasets**	**Common**	56	—	・When human annotation is required, did you consider measures for managing it (e.g. recording the annotation history)?	—	—	—	—	—
Data check	**3 Coverage of datasets**	**Common**	57	—	・Did you specifically check the quality of the test data set (e.g. there is no outlier or missing value, the label is correct, the person who labelled is clear, and the date, place, and history of acquisition is clear) in areas that are specifically related to safety, such as normal/abnormal judgment?	—	—	—	—	—
Data check	**3 Coverage of datasets**	**Common**	58	—	・Did you fully examine the validity of rule-based programming for data augmentation (e.g. creating data in which image data is made to be line symmetric to compensate for the lack of data)?	—	—	—	—	—

“Perspectives in the Field of Plant Safety” Checklist for Internal Quality Assurance

Situation of use	Internal quality requirements in “Machine Learning Quality Management Guideline 1st edition”				Perspectives in the field of plant safety	Related “Use case-specific perspectives” (see Section 3.3 of the text)				
	Internal quality axis	Required level of internal quality	Requirement No.	Internal quality requirements		Prediction of pipe wall thickness	Pipeline diagnostic imaging	Equipment deterioration diagnosis	Detection and diagnosis of early signs of abnormality	Optimization of operation
Data check	3 Coverage of datasets	Common	59	—	・Did you evaluate the adequacy of the augmented data? Did you evaluate whether the development assumptions were appropriate for the distribution and labeling of additional data obtained during operation?	—	—	—	—	—
Data check	4 Uniformity of datasets	LvE1	60	(Same as the previous section "Coverage of datasets" Lv1.)	—	—	—	—	—	—
Data check	4 Uniformity of datasets	LvE1	61	Consider the source and method of acquiring test datasets to ensure that no bias is found in application situations.	・Did you pay attention to the source and method of obtaining data sets at the plant?	—	—	—	—	—
Data check	4 Uniformity of datasets	LvE1	62	Extract samples without bias from original data for each case to ensure that no bias is found.	・When there is a bias in the amount of measured value data, did you consider using a simulator?	—	—	—	—	—
Data check	4 Uniformity of datasets	LvE1	63	Record activities carried out to prevent bias from entering.	—	—	—	—	—	—
Data check	4 Uniformity of datasets	LvE1	64	Check that there are sufficient training data and test data for each analyzed case in the training phase, validation phase, and so on.	—	—	—	—	—	—
Data check	4 Uniformity of datasets	LvE1	65	When sufficient training data cannot be acquired for any case, review and loose the coverage standards and record what should be checked individually by system integration tests in line with the original standards.	・When the amount of data in a specific range is not sufficient, have you considered the possibility that the classification and prediction accuracy of that range could become low?	・ When the amount of data in a specific range is not sufficient, recognize that the prediction accuracy of that range could become low.	・ When the amount of data in a specific range is not sufficient, keep in mind that the prediction accuracy of that range could become low.	・Recognize that if sufficient operation data for a certain state cannot be obtained, the accuracy of detecting deterioration deviating from that state may be reduced.	・ Recognize that if sufficient normal data for a certain range cannot be obtained, the accuracy of detecting abnormality within that range may be reduced.	—
Data check	4 Uniformity of datasets	LvE2	66	Same as Lv 2 in “Coverage of datasets” in the previous section.	—	—	—	—	—	—
Data check	4 Uniformity of datasets	LvE2	67	The following activities are carried out in addition to those listed in Lv 1.	—	—	—	—	—	—
Data check	4 Uniformity of datasets	LvE2	68	Grasp an approximate probability of occurrence for each attribute value or each case.	—	—	—	—	—	—
Data check	4 Uniformity of datasets	LvE2	69	Check if acquired data is not deviated from the distribution.	—	—	—	—	—	—
Data check	4 Uniformity of datasets	LvE2	70	Positive check other than acquisition methods should be made regarding the coverage of the data included in each case.	—	—	—	—	—	—
Data check	4 Uniformity of datasets	LvE2	71	For example, in each case, when there is any attribute not included in that case, extract the distribution related to attribute and check if there is no significant bias.	—	—	—	—	—	—
Data check	4 Uniformity of datasets	LvE2	72	However, assumed probabilities of occurrence are compared with the whole sets of assumed events.	—	—	—	—	—	—
Data check	4 Uniformity of datasets	LvS1	73	Regarding the amount of data for each case considered in L1 of the previous section, explicitly check if there is a sufficient amount of data for risk cases.	—	—	—	—	—	—
Data check	4 Uniformity of datasets	LvS1	74	When data of rare cases is insufficient for training, comparing the amount of the whole sets of training data with a probability of occurrence of rare cases, consider focusing on learning of rare cases. However, especially when Lv E2 is required, with prioritized, the impact of reduced	—	—	—	—	—	—
Data check	4 Uniformity of datasets	LvS2	75	In addition to what is listed in Lv S1, estimate and design in advance the amount of data of each case, based on the estimated probability of occurrence for each risk event / case.	—	—	—	—	—	—
Data check	4 Uniformity of datasets	Common	76	—	・Did plant engineers on site confirm that there was no bias in the test data set for each case?	—	—	—	—	—

“Perspectives in the Field of Plant Safety” Checklist for Internal Quality Assurance

Situation of use	Internal quality requirements in “Machine Learning Quality Management Guideline 1st edition”				Perspectives in the field of plant safety	Related “Use case-specific perspectives” (see Section 3.3 of the text)				
	Internal quality axis	Required level of internal quality	Requirement No.	Internal quality requirements		Prediction of pipe wall thickness	Pipeline diagnostic imaging	Equipment deterioration diagnosis	Detection and diagnosis of early signs of abnormality	Optimization of operation
Data check	**4 Uniformity of datasets**	**Common**	77	—	・Did you consider the possibility that data characteristics, including the frequency of data generation, may change due to equipment switching, maintenance, etc.?	—	—	—	—	—
Data check	**4 Uniformity of datasets**	**Common**	78	—	・Did you pay attention to linking with the plant operation data as the data collection interval could be several months or years?	—	—	—	—	—
Data check	**4 Uniformity of datasets**	**Common**	79	—	・Did you confirm that the data within the assumed range was obtained without bias?	・ Pay attention to ensure that the amount of data for each range of data to be covered by the abovementioned attributes is sufficient.	・ Pay attention to ensure that the amount of data for each range of data to be covered by the abovementioned attributes is sufficient.	・ Obtain operation data in various states without bias assumed as "No deterioration"	・ Obtain data without bias in various ranges (e.g. daytime/night, steady/unsteady, seasonal differences) assumed as normal data.	—
Data check	**4 Uniformity of datasets**	**Common**	80	—	・When the data has biases, have you assessed selection bias, information bias, and confounding problems and risks? Have you removed or corrected outliers or missing values based on the acceptance and elimination policy?	—	—	—	—	—
Learning	**5 Correctness of the trained model**	**Lv1**	81	Prepare a test dataset by taking into account the coverage of data and the past experiences, e.g., obtained in the proof of concept (POC) stage.	—	—	—	—	—	—
Learning	**5 Correctness of the trained model**	**Lv1**	82	Decide and explain how to ensure the quality of the test dataset, such as excluding outliers and modifying incorrect labels.	—	—	—	—	—	—
Learning	**5 Correctness of the trained model**	**Lv1**	83	Prepare a training dataset in an analogous way to the test dataset. Note that the training dataset may not necessarily follow the same distribution as the test dataset.	—	—	—	—	—	—
Learning	**5 Correctness of the trained model**	**Lv1**	84	Decide and record how to deal with the incorrect behavior of a trained model (e.g., false negative/false positive in the test) before the validation phase.	・Did you consider the allowable level of misjudgment based on the assumed use on site in the field of plant safety?	—	—	—	・In this case, a certain amount of false detection can be tolerated, but since the amount and types of abnormal data that can be used for the test are limited, it is preferable for false detection rate to be as close to zero as possible.	・ In the case of "Assumption that no safety function is required from the presentation of operating parameters by ML components (when operating parameters are presented that exceed the assumed specifications of the equipment, they are not reflected in the operation by external safety mechanisms and operator judgment)," it is not required to consider "Keeping the permissible level of output of parameters leading to dangerous operation as close to zero as possible." Based on the SIL assessment, etc., determine the required level for the ML components by appropriately assigningthe safety functions to the ML components and other systems.
Learning	**5 Correctness of the trained model**	**Lv1**	85	If a machine learning component is required to satisfy fairness, decide and record the methods and criteria to evaluate fairness before the validation phase.	・("Fairness" is outside the scope of the Guidelines)	—	—	—	—	—
Learning	**5 Correctness of the trained model**	**Lv2**	86	All the requirements listed in Lv 1.	—	—	—	—	—	—
Learning	**5 Correctness of the trained model**	**Lv2**	87	Provide certain evidence to show the correctness of labels in the training, validation, and test datasets.	—	—	—	—	—	—
Learning	**5 Correctness of the trained model**	**Lv2**	88	Decide and explain methods and criteria to validate the trained model (e.g., accuracy and its threshold) before the validation phase.	—	—	—	—	—	—
Learning	**5 Correctness of the trained model**	**Lv2**	89	Test the trained model using the given test dataset and additional test data (e.g., generated by data augmentation techniques).	—	—	—	—	—	—
Learning	**5 Correctness of the trained model**	**Lv2**	90	If possible, analyze internal information on the trained model (e.g., the neuron coverage to evaluate the adequacy of testing).	—	—	—	—	—	—
Learning	**5 Correctness of the trained model**	**Lv3**	91	All the requirements listed in Lv 2.	—	—	—	—	—	—

“Perspectives in the Field of Plant Safety” Checklist for Internal Quality Assurance

Situation of use	Internal quality requirements in “Machine Learning Quality Management Guideline 1st edition”				Perspectives in the field of plant safety	Related “Use case-specific perspectives” (see Section 3.3 of the text)				
	Internal quality axis	Required level of internal quality	Requirement No.	Internal quality requirements		Prediction of pipe wall thickness	Pipeline diagnostic imaging	Equipment deterioration diagnosis	Detection and diagnosis of early signs of abnormality	Optimization of operation
Learning	5 Correctness of the trained model	Lv3	92	Perform the validation/testing of the whole system (e.g., integration tests), especially by focusing on risky cases.	—	—	—	—	—	—
Learning	5 Correctness of the trained model	Lv3	93	Reflect in the test plan the measures to be taken for the requirements of ML components during product-level testing, centering on the high-risk cases in particular (considered in 2 Coverage for distinguished problem cases), and monitor and check the coverage thereof.	—	—	—	—	—	—
Learning	5 Correctness of the trained model	Common	94	—	・Have you agreed with the concerned parties (e.g. plant operator, engineering company) on the accuracy evaluation criteria for the model?	—	—	—	—	—
Learning	5 Correctness of the trained model	Common	95	—	・Did you test the quality before the start of operation even when using reinforcement learning?	—	—	—	—	・ Even if reinforcement learning is used, meet the requirements of "Correctness of the trained model" by conducting tests before the start of operation.
Learning	5 Correctness of the trained model	Common	96	—	・Did you consider that the correct value that should be given depends on the problem, such as the label for identification problems and the value for regression problems?	—	—	—	—	—
Learning	5 Correctness of the trained model	Common	97	—	・Are data for training and testing used for cross-validation and generalization performance, etc. and validation data separated and managed independently? Are data for re-learning and additional learning managed in the same manner?	—	—	—	—	—
Learning	5 Correctness of the trained model	Common	98	—	・Have the accuracy and the residual error of loss function sufficiently converged after learning? Have the precision, recall, and F-measure reached the target level?	—	—	—	—	—
Learning	5 Correctness of the trained model	Common	99	—	・Aren’t the accuracy and the residual error of loss function of the learning/re-learning process showing abnormal changes?	—	—	—	—	—
Learning	5 Correctness of the trained model	Common	100	—	・Have you clarified the basis for selecting the AI algorithm and distillation or non-distillation, and the basis for setting hyperparameters? Have you explained and agreed on the rationale for algorithm selection among users and vendors?	—	—	—	—	—
Learning	5 Correctness of the trained model	Common	101	—	・Did you check the need to reduce the size of the learning data set or learned models? At that time, did you consider how much performance deterioration would be tolerated?	—	—	—	—	—
Learning	6 Stability of the trained model	Lv1	102	Record the concrete techniques applied to improve the generalization performance of a trained model (e.g., cross validation and regularization are widely used to prevent overfitting to the training data).	—	—	—	—	—	—
Learning	6 Stability of the trained model	Lv1	103	In Lv1, it is recommended to apply techniques such as cross-validation and regularization, which are widely used to prevent overlearning.	—	—	—	—	—	—
Learning	6 Stability of the trained model	Lv2	104	Record the evaluation results of stability by using neighboring data.	—	—	—	—	—	—

“Perspectives in the Field of Plant Safety” Checklist for Internal Quality Assurance

Situation of use	Internal quality requirements in “Machine Learning Quality Management Guideline 1st edition”				Perspectives in the field of plant safety	Related “Use case-specific perspectives” (see Section 3.3 of the text)				
	Internal quality axis	Required level of internal quality	Requirement No.	Internal quality requirements		Prediction of pipe wall thickness	Pipeline diagnostic imaging	Equipment deterioration diagnosis	Detection and diagnosis of early signs of abnormality	Optimization of operation
Learning	6 Stability of the trained model	Lv2	105	In the evaluation of stability at Lv2, it is required to use some synthetic data obtained by adding a small amount of noise to the learning datasets. In particular, it is recommended to apply techniques to prevent attacks based on adversarial examples; e.g., robustness evaluation using adversarial examples, adversarial training to train a robust model, and dynamic detection of adversarial examples. Currently, these new methods are still being studied and developed in academic research, but might be applied to system development more effectively in the future.	—	—	—	—	—	—
Learning	6 Stability of the trained model	Lv3	106	Provide formal guarantee to the stability for neighboring data.	—	—	—	—	—	—
Learning	6 Stability of the trained model	Lv3	107	At Lv 3, it is required to formally guarantee a certain level of stability for neighboring data. For example, methods for certifying adversarial robustness have been studied recently and might be used in system development in the future.	—	—	—	—	—	—
Learning	6 Stability of the trained model	Common	108	—	・Did you consider the need to pay attention to the stability of near-field data of learning data sets, especially in chemical plants, due to the high uncertainty of generated data?	—	—	—	—	—
Learning	6 Stability of the trained model	Common	109	—	・Did you consider that paying attention to stability is necessary for plants with new products added frequently, as the data generated lies mostly near the training dataset?	—	*This use case does not apply when data on product characteristics (type, contents, etc.) are not utilized.	—	—	—
Learning	6 Stability of the trained model	Common	110	—	・Have you discussed and aligned on what generalization performance measurements are appropriate?	—	—	—	—	—
Learning	6 Stability of the trained model	Common	111	—	・Have you clearly defined the target value of generalization performance? Isn’t the generalization performance of the AI model after learning significantly deteriorated compared with the accuracy during learning?	—	—	—	—	—
Learning	6 Stability of the trained model	Common	112	—	・Have you decided how to measure generalization performance? When using cross-validation, do you secure a variation of the learning data set to be used?	—	—	—	—	—
Learning	6 Stability of the trained model	Common	113	—	・Do you identify noise candidates that affect AI? Specifically, do you select error factors and analyze their effects? Is there a case where noise candidates cause a significant deterioration in the performance of the AI model?	—	—	—	—	—
Implementation and operation	7 Reliability of underlying software systems	Lv1	114	Select reliable software used for the machine learning system, and record the process of this selection.	・When using a simulator, have you checked the track record and recorded the selection process?	—	—	—	—	—
Implementation and operation	7 Reliability of underlying software systems	Lv1	115	Monitor the system’s operation to check and update the selected software.	・When using a simulator, is it possible to monitor and correct defects?	—	—	—	—	—

“Perspectives in the Field of Plant Safety” Checklist for Internal Quality Assurance

Situation of use	Internal quality requirements in “Machine Learning Quality Management Guideline 1st edition” Internal quality axis	Required level of internal quality	Requirement No.	Internal quality requirements	Perspectives in the field of plant safety	Related “Use case-specific perspectives” (see Section 3.3 of the text) Prediction of pipe wall thickness	Pipeline diagnostic imaging	Equipment deterioration diagnosis	Detection and diagnosis of early signs of abnormality	Optimization of operation
Implementation and operation	7 Reliability of underlying software systems	Lv1	116	Examine in advance the impact of differences between the environment in the training/test phases and the environment in the actual operation phase.	—	—	—	—	—	—
Implementation and operation	7 Reliability of underlying software systems	Lv2	117	Evaluate the reliability of the software used for the system by testing.	・When using a simulator, have you self-evaluated its reliability through inspections, experiments, etc.?	—	—	—	—	—
Implementation and operation	7 Reliability of underlying software systems	Lv2	118	If possible, use software whose reliability is SIL 1 or equivalent.	—	—	—	—	—	—
Implementation and operation	7 Reliability of underlying software systems	Lv2	119	Prepare for the maintenance of software during its operation.	・When using a simulator, do you have a maintenance system in place?	—	—	—	—	—
Implementation and operation	7 Reliability of underlying software systems	Lv2	120	In the validation and test phases, conduct validation tests in an environment that simulates the environment used in the actual operation phase. Alternatively, validate the consistency of operations of the trained model between in the test phase and the actual operation phase.	—	—	—	—	—	—
Implementation and operation	7 Reliability of underlying software systems	Lv3	121	Check the quality of software for SIL 1 (or a higher SIL level when required by the system).	—	—	—	—	—	—
Implementation and operation	7 Reliability of underlying software systems	Lv3	122	Perform testing or formal verification of the behaviors of the trained model in an actual environment.	—	—	—	—	—	—
Implementation and operation	7 Reliability of underlying software systems	Lv3	123	Check the consistency of those models and operations in an actual environment in any stage after integration tests.	—	—	—	—	—	—
Implementation and operation	7 Reliability of underlying software systems	Common	124	—	・Did you give consideration to keep the computational effort of ML components at an appropriate level for the following cases? - When computational resources are limited due to a special environment or device (when calculating with an edge device, a special PC, etc.) - When other processes are running on the same device and they cannot be affected, etc.	—	—	—	—	—
Implementation and operation	7 Reliability of underlying software systems	Common	125	—	・When assessing the system, is unit testing against external libraries or system testing including external libraries performed?	—	—	—	—	—
Implementation and operation	7 Reliability of underlying software systems	Common	126	—	・Have you agreed on the specific scope of responsibility for defects with your library suppliers?	—	—	—	—	—
Implementation and operation	7 Reliability of underlying software systems	Common	127	—	・Do you use software considering the update frequency and support period of various types of software such as OS and OSS? Have you decided what to do with software updates and what to do when support ends?	—	—	—	—	—
Implementation and operation	7 Reliability of underlying software systems	Common	128	—	・Do you update systems along with software updates such as OS and OSS, especially when security updates are available?	—	—	—	—	—
Implementation and operation	8 Maintainability of qualities in use	Lv1	129	Examine in advance how to respond to notable system quality deterioration caused by changes in external environment.	・Did you extract "Changes in the external environment" in the plant field? -Plant repairs, aging, changes in operating conditions, etc.	—	—	—	—	—

“Perspectives in the Field of Plant Safety” Checklist for Internal Quality Assurance

Situation of use	Internal quality requirements in “Machine Learning Quality Management Guideline 1st edition”				Perspectives in the field of plant safety	Related “Use case-specific perspectives” (see Section 3.3 of the text)				
	Internal quality axis	Required level of internal quality	Requirement No.	Internal quality requirements		Prediction of pipe wall thickness	Pipeline diagnostic imaging	Equipment deterioration diagnosis	Detection and diagnosis of early signs of abnormality	Optimization of operation
Implementation and operation	**8 Maintainability of qualities in use**	**Lv1**	130	In the case where online learning is given, examine in advance the impact of unexpected quality deterioration and take measures from the system side such as the limitation of operation range if necessary.	—	—	—	—	—	—
Implementation and operation	**8 Maintainability of qualities in use**	**Lv1**	131	When additional learning is given off-line, quality management in line with the previous seven paragraph should be introduced.	—	—	—	—	—	—
Implementation and operation	**8 Maintainability of qualities in use**	**Lv2**	132	Monitor system quality deterioration and misjudgments by comparing with operation results within the range permitted by system use. It is necessary to sufficiently examine factors other than product quality such as privacy at the time of monitoring.	・Did you extract "Factors other than product quality" in the plant field? - Degree of data disclosure from plant operators to AI vendors, etc.	—	—	—	—	—
Implementation and operation	**8 Maintainability of qualities in use**	**Lv2**	133	When online learning is given, regularly monitor additional learning results by any method. When any deviation from the requirements for performance is found as a result of monitoring, an immediate handling should be taken.	—	—	—	—	—	—
Implementation and operation	**8 Maintainability of qualities in use**	**Lv2**	134	When additional learning is given off-line, conduct “regression tests on quality deterioration” with test datasets used in the system development stage to check if the quality has been maintained prior to updates. Update test datasets using the same method used in the system development stage where necessary.	—	—	—	—	—	—
Implementation and operation	**8 Maintainability of qualities in use**	**Lv3**	135	Make sure to establish measures for monitoring system quality, including an operational system, compatible to privacy.	—	—	—	—	—	—
Implementation and operation	**8 Maintainability of qualities in use**	**Lv3**	136	When online learning is given, before the results of additional learning are reflected on systems, implement a mechanism to check quality to some extent within those systems so that updates are suspended if it becomes impossible to ignore unexpected quality deterioration. Make sure to ensure measures for making updates and modifications off-line.	—	—	—	—	—	—
Implementation and operation	**8 Maintainability of qualities in use**	**Lv3**	137	When additional learning is given off-line, the quality should be managed using data collected from operation, test datasets used for the initial system building and test datasets updated on a regular basis using the same method.	—	—	—	—	—	—
Implementation and operation	**8 Maintainability of qualities in use**	**Common**	138	—	・Do you have a system in place for re-learning and accuracy re-verification of models to ensure the Maintainability of qualities in use?	—	—	—	—	—
Implementation and operation	**8 Maintainability of qualities in use**	**Common**	139	—	・Did you check with plant engineers on site whether the quality monitoring system for operation is appropriate?	—	—	—	—	—
Implementation and operation	**8 Maintainability of qualities in use**	**Common**	140	—	・Did you pay attention to the possibility that the characteristics of the generated data may change due to maintenance, such as replacement or adjustment of parts or modification of equipment?	—	—	・ Note that when the type of the equipment is replaced, it may be necessary to take measures such as re-learning and switching of the learning model.	—	—

“Perspectives in the Field of Plant Safety” Checklist for Internal Quality Assurance

Situation of use	Internal quality requirements in “Machine Learning Quality Management Guideline 1st edition”				Perspectives in the field of plant safety	Related “Use case-specific perspectives” (see Section 3.3 of the text)				
	Internal quality axis	Required level of internal quality	Requirement No.	Internal quality requirements		Prediction of pipe wall thickness	Pipeline diagnostic imaging	Equipment deterioration diagnosis	Detection and diagnosis of early signs of abnormality	Optimization of operation
Implementation and operation	8 Maintainability of qualities in use	Common	141	—	・Did you consider a monitoring system for understanding changes in data characteristics?	—	—	—	—	・ Monitor the evaluation of optimum values on a regular and continuous basis to check for abnormalities.
Implementation and operation	8 Maintainability of qualities in use	Common	142	—	・Did you confirm that re-learning and accuracy re-verification of the model are necessary because it is likely that the characteristics of the generated data will change if the component values of the product change?	—	*This use case does not apply when data on product characteristics (type, contents, etc.) are not utilized.	・ Ensure that quality is maintained when entering product component values of changing product component values.	—	—
Implementation and operation	8 Maintainability of qualities in use	Common	143	—	・Did you confirm that re-learning and accuracy re-verification of the model are necessary when the assumptions surrounding the equipment or operating procedures change?	—	—	・ Check if there are any deviations from assumptions such as environmental factors that were assumed at the beginning, including not only the target equipment itself but also surrounding conditions.	・ Since changes in the external environment (e.g. sunlight, wind direction) have a particularly large impact on ML components in chemical plants, pay attention to changes that affect the external environment of the target equipment, even if they are not changes of the equipment, such as removal or modification of adjacent equipment.	・ When a change is made to equipment or operating procedures, update the model because it will affect the output of ML components.
Implementation and operation	8 Maintainability of qualities in use	Common	144	—	・Did you consider periodically verifying the accuracy of data collected during operation?	—	・ Verify accuracy using an image photographed during the operation phase. Record the results of visual inspections performed upon AI's decision, and compare these results to the results of aforementioned verification. If the judgment accuracy is low, perform a thorough check on input image and model used in the verification.	—	—	—
Implementation and operation	8 Maintainability of qualities in use	Common	145	—	・When aging is expected to progress quickly, did you design the frequency of accuracy verification and tuning of the learning model accordingly?	—	—	—	・ Expect aging to progress depending on the production load of the target equipment and design the frequency of accuracy verification and tuning of the learning model accordingly.	—
Implementation and operation	8 Maintainability of qualities in use	Common	146	—	・Did you confirm that accuracy verification and tuning of the learning model are necessary each time when the target equipment is repaired on a large scale (not aging)?	—	—	—	・ Accuracy verification and tuning of the learning model are necessary each time when the target equipment is repaired on a large scale (not aging).	—
Implementation and operation	8 Maintainability of qualities in use	Common	147	—	・Did you check the accuracy, etc. of ML components based on a comparison with the results of conventional methods that do not rely on machine learning or actual results?	・ Based on the judgment of whether or not replacement is necessary using the existing method, the actual condition of piping at the time of replacement, and others, verify the actual accuracy and the presence of oversight.	—	—	—	—
Implementation and operation	8 Maintainability of qualities in use	Common	148	—	In particular, when accuracy needs to be maintained, did you prepare for change management after the start of operation by organizing assumptions and training data when constructing a model in advance?	・ As maintaining accuracy is vital in this case, it is crucial to keep extensive records of background information on model construction and types of training data. Consult these records every time changes are to be made after operation begins.	—	—	—	—
Implementation and operation	8 Maintainability of qualities in use	Common	149	—	Did you consider limiting the output range of ML components as needed?	—	—	—	—	・ As stability may be impaired if operating conditions are pursued to the limit with respect to the optimization target, take measures such as limiting the output range of ML components.
Implementation and operation	8 Maintainability of qualities in use	Common	150	—	Did you check the assumptions (e.g. assumed output range, equipment condition setting) to check the quality during operation?	—	—	—	—	・ Confirm that the equipment is operated within the expected interpolation range of the assumed raw material (e.g. crude oil type). ・ Confirm the output quality of ML components by considering various conditions (e.g. initial period of/end of reaction, operating conditions, raw materials, quality requirements, allowable time for startup and shutdown) of equipment under operation.
Implementation and operation	8 Maintainability of qualities in use	Common	151	—	・Do you have a system in place to record data that can only be collected in actual operation? Did you secure data for errors and diversity found during operation?	—	—	—	—	—

“Perspectives in the Field of Plant Safety” Checklist for Internal Quality Assurance

Situation of use	Internal quality requirements in “Machine Learning Quality Management Guideline 1st edition”				Perspectives in the field of plant safety	Related “Use case-specific perspectives” (see Section 3.3 of the text)				
	Internal quality axis	Required level of internal quality	Requirement No.	Internal quality requirements		Prediction of pipe wall thickness	Pipeline diagnostic imaging	Equipment deterioration diagnosis	Detection and diagnosis of early signs of abnormality	Optimization of operation
Implementation and operation	8 Maintainability of qualities in use	Common	152	—	・Do you identify any bias that is different from when the operational data was introduced, and analyze the background? Are the grounds and measures for removing and correcting outliers and missing values based on the acceptance and elimination policy? Are you prepared to maintain the system?	—	—	—	—	—
Implementation and operation	8 Maintainability of qualities in use	Common	153	—	・Do you monitor the quality of input data, such as monitoring whether the data to update the model is out of the assumed data range?	—	—	—	—	—
Implementation and operation	8 Maintainability of qualities in use	Common	154	—	・After the start of operation, do you extract factors that affect performance and set performance targets with margins? Does the design allow humans or an AI system to judge the detection of performance deterioration?	—	—	—	—	—
Implementation and operation	8 Maintainability of qualities in use	Common	155	—	・Did you decide on a validation method that is still applicable when the training dataset increases in variation?	—	—	—	—	—
Implementation and operation	8 Maintainability of qualities in use	Common	156	—	・As a result of re-learning due to changes in the characteristics of the training data, addition of output, etc., is the performance deterioration before re-training acceptable?	—	—	—	—	—
Implementation and operation	8 Maintainability of qualities in use	Common	157	—	・When updating the AI model automatically rather than manually, can you fully check that characteristic and performance changes of the model are acceptable?	—	—	—	—	—
Implementation and operation	8 Maintainability of qualities in use	Common	158	—	・For the data fed back to learning, is it possible to prevent the inclusion of inappropriate data (e.g. input data in operations obtained from a population different from the learning data in learning, outliers of input data) that can lead to performance degradation? Or, is there a mechanism to eliminate inappropriate data before learning?	—	—	—	—	—
Implementation and operation	8 Maintainability of qualities in use	Common	159	—	・Regarding the input data used for inference in operation, is there a mechanism in place to eliminate abnormality that can lead to an abnormal behavior or inappropriate data (e.g. input data in operation obtained from a population different from the learning data in learning, outliers of input data)?	—	—	—	—	—
Implementation and operation	8 Maintainability of qualities in use	Common	160	—	・Did you consider a method of delivering a model that has completed re-learning?	—	—	—	—	—
Implementation and operation	8 Maintainability of qualities in use	Common	161	—	・Do you have a mechanism in place to roll back quickly when an abnormality occurs in a released AI program?	—	—	—	—	—

[Annex]

List of Members, etc. for the Working Group on AI Reliability Assessment in Plants

Chairperson (honorific title omitted)

YAMASHITA Yoshiyuki	Professor, Department of Chemical Engineering, Tokyo University of Agriculture and Technology

Members (in Japanese syllabary order, honorific titles omitted)

AOYAMA Atsushi	Professor, Graduate School of Technology Management, Ritsumeikan University
ISHIKAWA Fuyuki	Associate Professor, Information Systems Architecture Science Research Division, National Institute of Informatics
ESAKI Norio	General Manager, Advanced Production Technology Promotion Division, Production & Technology Center, Mitsui Chemicals, Inc.
OIWA Yutaka	Leader, Software Quality Assurance Research Team, Cyber Physical Security Research Center, National Institute of Advanced Industrial Science and Technology (AIST)
OONO Takuya	Associate Executive Officer, Process Technology Division, Oil & Gas Project Company, JGC Corporation
OKADA Yoshiaki	Executive Specialist, Corporate Business Development Division, NEC Corporation
OYAMADA Kenji	Deputy General Manager, High Pressure Gas Safety Department, The High Pressure Gas Safety Institute of Japan (KHK)
KUNITŌ Yōichi	General Manager, Safety Engineering Center , Production & Technology Department, Mitsubishi Chemical Corporation
KOBUCHI Keiichirō	AI Business Senior Manager, Information Technology Center, IA Product & Service Business Headquarters, Yokogawa Electric Corporation
KONDŌ Akihiro	Industrial PA Solution Department, Solution & Service Division, Industrial & Distribution Business Unit, Hitachi, Ltd.
SAITŌ Tsuyoshi	Senior Researcher, Mechanical System Safety Research Group, National Institute of Occupational Safety & Health, Japan
SHOKUDA Hideki	Senior Manager, System Innovation Technology Office, Technology & Engineering Center, Idemitsu Kosan Co.,Ltd.
NISHI Haruki	Divison Director, Fire Disaster Investigation Division, National Research Institute of Fire and Disaster
HAMADA Satoshi	Environment and Safety Office, Responsible Care Department, Showa Denko K.K.
HARADA Noriaki	General Manager Digital Innovation Center, Corporate Production Technology, Asahi Kasei Corporation
FUJII Tatsuya	Facilities Management Group Manager, Engineering Department, ENEOS Corporation
MASUTANI Masataka	General Manager, Production Technology Department, JSR Corporation
MARUYAMA Hiroshi	PFN Fellow, Preferred Networks, Inc.
YASUI Takehito	Senior Specialist, Digital Planning and Marketing Department, Digital Transformation Division, Chiyoda Corporation

Observers (honorific titles omitted)

Petroleum Association of Japan
Japan Petrochemical Industry Association
Japan Chemical Industry Association
Engineering Advancement Association of Japan
Japan Electric Measuring Instruments Manufacturers' Association
Japan Deep Learning Association
Japan Association of Maintenance and Service Contractors
Research Institute of Economy, Trade and Industry
Fire and Safety Section, Disaster Prevention Department, Living Safety and Disaster Prevention Bureau, Kanagawa Prefecture
Manufacturing Industry Promotion Division, Employment and Economy Department, Mie Prefecture
Commerce and Industry Division, Commerce, Industry, Agriculture and Fisheries Department, Yokkaichi City

Extraordinary Disaster Management Office, Fire and Disaster Management Agency, Ministry of Internal Affairs and Communications
Dangerous Goods Safety Office, Fire and Disaster Management Agency, Ministry of Internal Affairs and Communications
Safety Division, Industrial Safety and Health Department, Labour Standards Bureau, Ministry of Health, Labour and Welfare
Chemical Hazards Control Division, Industrial Safety and Health Department, Labour Standards Bureau, Ministry of Health, Labour and Welfare
Material Industries Division, Manufacturing Industries Bureau, Ministry of Economy, Trade and Industry
Information Economy Division, Commerce and Information Policy Bureau, Ministry of Economy, Trade and Industry
High Pressure Gas Safety Office, Industrial and Product Safety Policy Group, Ministry of Economy, Trade and Industry
Petroleum Refining and Reserve Division, Natural Resources and Fuel Department, Agency for Natural Resources and Energy, Ministry of Economy, Trade and Industry

Secretariat

Mitsubishi Research Institute, Inc.

Guidelines on Assessment of AI Reliability in the Field of Plant Safety
Second Edition

March 2021

The Liaison Council of Three Ministries on Disaster Prevention of Petroleum Complexes
(Ministry of Economy, Trade and Industry; Fire and Disaster Management Agency; and Ministry of Health, Labour and Welfare)

"Working Group on AI Reliability Assessment in Plants"
Secretariat: Mitsubishi Research Institute