当翻訳は仮訳であり、正確には原文を参照してください。
Please refer to the original text for accuracy

仮訳
Provisional Translation

Ministry of Economy, Trade and Industry

Guidance on the Interpretation and Application of Civil Liability in the Utilization and Application of AI

[Version 1.0]

April 2026

Contents

Chapter 1 Introduction

1.1 Purpose

Since the advent of generative AI, the utilization of AI services in the business sector has been increasing year by year. Recent years have seen the progress in the development and social implementation of sophisticated systems, such as IoT products and robotics equipped with AI. The sophistication of systems has increased needs for large-scale, high-quality data for the development and operation of systems, which has brought expectations for the creation of new added values and for the solution of social problems through data linkage across the boundaries between businesses.

Amid this increase in the effect of the advanced use of AI and data not only on cyberspace but also on physical space, the national government, with the intention of reducing social risks brought on AI and promoting the innovation and utilization of AI, established the AI Guidelines for Business[1] (Ver. 1.0 published in April 2024, revised to Ver. 1.2 in March 2026). As for matters to be taken into account in the context of contracts, the government established the Contract Guidelines on Utilization of AI and Data[2] (published in June 2018, revised in December 2019; hereinafter referred to as the "Contract Guidelines") and the Contract Checklist for AI Use and Development[3] (published in February 2025; hereinafter referred to as the "Contract Checklist"), among others. In addition, the Interim Report[4] on the modality of AI governance was issued by the AI Strategy Council and the AI System Study Group in February 2025, followed by the enactment of the Act on Promotion of Research and Development, and Utilization of AI-related Technology (Act No. 53 of 2025) in May 2025, which sets forth the function of the national government as the control tower and the responsibilities of related parties.

Looking at actions taken in foreign countries, we find that in Europe, the European Parliament adopted the Report with Recommendations to the Commission on Civil Law Rules on Robotics[5] in January 2017 regarding civil liability of robots and AI, followed by the adoption of the resolution

[1] Ministry of Internal Affairs and Communications and Ministry of Economy, Trade and Industry. *AI Jigyosya Guidelines (Dai 1.2-han) [AI Guidelines for Business (Ver. 1.2)]*. March 31, 2026. https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20260331_1.pdf (last accessed on December 31, 2026).

[2] Ministry of Economy, Trade and Industry. *AI・Deta no Riyo ni Kansuru Guidelines 1.1-han [Contract Guidelines on Utilization of AI and Data Ver. 1.1]*. December 2019. https://www.meti.go.jp/policy/mono_info_service/connected_industries/sharing_and_utilization/20200619001.pdf (last accessed on December 25, 2025).

[3] Ministry of Economy, Trade and Industry. *AI no Riyo・Kaihatsu ni Kansuru Checklist [Contract Checklist for AI Use and Development].* February 2025. https://www.meti.go.jp/policy/mono_info_service/connected_industries/sharing_and_utilization/20250218003-ar.pdf (last accessed on December 25, 2025).

[4] AI Strategy Council and AI System Study Group. *Chukan Torimatome [Interim Report]*. February 4, 2025. https://www8.cao.go.jp/cstp/ai/interim_report.pdf (last accessed on December 25, 2025).

[5] EU. REPORT with recommendations to the Commission on Civil Law Rules on Robotics. January 27, 2017. https://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.html (last accessed on December 25, 2025).

on a civil liability regime for AI[6] in October 2020, which recommends applying the Product Liability Directive to AI. Subsequently, in December 2024, the revised Product Liability Directive[7] came into effect, which resulted in software and certain other intangible items being included in "products" covered by this directive. As a result, product liability now applied to AI. In addition, August 2024 saw the enforcement of the EU AI Act, which is a comprehensive advance regulation of AI. As of the date of writing of this Guidance, the provisions relating to prohibited AI systems and the like and those relating to general-purpose AI models and the like have been put into effect. Furthermore, some of the provisions relating to high-risk AI will be enforced in August 2026, with all other regulations, including those relating to high-risk AI, to be enforced in August 2027.[8] In the U.S., on the other hand, the Transparency in Frontier Artificial Intelligence Act was enacted in California in September 2025. AI-related lawsuits that have been filed include those regarding biases in AI recruiting tools and those regarding traffic accidents caused by autonomous driving. Discussions are proceeding as to where liability should fall, through the accumulation of practical experiences.

Japan has taken a policy to deal with risks posed by AI by combining the existing legal system and soft law together, and to consider additional legal regulations only on risks that cannot be expected to be dealt with voluntarily by businesses.[9] These and other circumstances indicate that the existing civil liability rules will play a key role in resolving cases involving AI-related damage or the like. On the other hand, there has been the opinion that one factor that has discouraged the development and introduction of AI is the lack of accumulation of judicial precedents and the lack of a unified view as to how civil liability should be interpreted and applied in the event of rights infringement or damage upon realization of the risks, in light of the autonomy and black box nature of AI. At meetings of the Working Group on AI Guidelines for Business, some opinions were expressed that pointed out the necessity to discuss the demarcation of liability in the development, provision, and utilization of AI.[10]

In light of the trends in foreign countries and the challenges found in Japan, the Ministry of Economy, Trade and Industry recently held meetings of the Study Group on Civil Liability in the Utilization of AI (hereinafter referred to as the "Study Group"), at which issues and positions in the interpretation and application of civil liability were organized mainly from the viewpoint of tort law, among others, based on hypothetical basic cases in which AI-powered services or systems

---

[6] EU. Civil liability regime for artificial intelligence. October 20, 2020. https://oeil.europarl.europa.eu/oeil/en/procedure-file?reference=2020/2014(INL) (last accessed on December 25, 2025).

[7] EU member states are required to reflect this directive into their national law by December 2026.

[8] The EU Digital Omnibus Regulation Proposal indicates a plan to postpone the application of certain provisions relating to high-risk AI until December 2027 at latest and all other provisions, including regulations on high-risk AI, until August 2028 at latest (EU. Digital Omnibus on AI Regulation Proposal. November 19, 2025. https://digital-strategy.ec.europa.eu/en/library/digital-omnibus-ai-regulation-proposal (last accessed on December 25, 2025)).

[9] Id. note 4. AI Strategy Council and AI System Study Group. pp. 8-10.

[10] Conference Toward AI Network Society (30th meeting) and AI Governance Study Group (26th meeting), Ministry of Internal Affairs and Communications. March 7, 2025. https://www.soumu.go.jp/main_content/001000976.pdf (last accessed in December 25, 2025).

contribute to an accident. The Study Group mainly addressed tort law, and did not define a precise delineation of liability between parties that takes their contractual relationships and such other factors also into account. Still, tort law provides basic default rules as to what liability is imposed on each party in situations not covered by a contract or other explicit rules. Suggestions from the viewpoint of tort law can also provide considerations in the process of further defining where liability lies in contracts between the parties or in guidelines or other documents.

This Guidance provides a summary of the discussions undertaken by the Study Group, and intends to contribute to the promotion of utilization of AI and to the smooth resolution of cases of damage caused by AI, by increasing foreseeability by parties involved in the development, provision, and utilization of AI through indicating potential directions in which existing laws can be applied, focusing on tort law issues that may arise in settings where AI is utilized. With this intention, the commentary of each potential case: mainly explains the directions in which existing laws should be interpreted and applied, such as what factors increase or decrease the possibility of a party being held liable; extracts views that would be useful in resolving as wide variety as possible of cases; and attempts to provide the assumptions for and scope of validity of each of these views. It should be noted that it is the courts that are authorized to make final judgments on how existing laws apply to specific individual cases, and that there is no guarantee that the views described in this Guidance will be adopted by the courts without change. However, it generally takes a long time before legal interpretation and application are definitely established through the accumulation of judicial precedents. We hope that this Guidance, which presents, based on discussions by experts, most reasonable views possible and issues over which opinions are divided at this point in time, will serve as a material for further discussions on legal interpretation and will eventually be of some help in making new rules.

Due to its nature, this Guidance contains highly legally or technologically technical descriptions. Its expected readers are businesses engaged in the development, provision, or utilization of AI, parties and lawyers working to resolve disputes involving the utilization of AI, and a wide range of such other entities.

1.2 Scope of This Guidance and Reservations

Please keep in mind the following reservations when using this Guidance:

- In light of the fact that the concept of liability in the settings where AI is utilized is not necessarily clear as described in 1.1 above, this Guidance presents views on the interpretation and application of civil liability based on existing tort law, and does not establish new rules.
- This Guidance is based on the technological trends and on the details of the legal system at the time of its writing. Once these assumptions change, the discussions presented in this Guidance may no longer be valid. This Guidance may be revised if necessary due to changes in the technological trends or the legal system.

- The views presented on the hypothetical cases described in Chapters 3 and 4 below are based on the respective relevant factual assumptions. In cases where the actual AI service works differently from the relevant assumptions or is used in settings that are different from the relevant assumptions, or where rights are infringed in a manner different from the assumptions, the discussions in this Guidance may not be valid. Each of the hypothetical cases was set for the purpose of organizing basic legal relations and does not provide comprehensive viewpoints to resolve actual cases.
- All hypothetical cases discussed by the Study Group are fictitious and are not related to any real AI services, businesses, individuals, etc.
- For the purpose of clarifying basic legal relationships between the parties under tort law, each of the hypothetical cases are based on relatively simple factual assumptions. For instance, the Study Group did not discuss cases where the AI developer, AI provider, and AI business user are all separate entities or where so-called “open source AI” is used.
- As described in 1.1 above, it is the courts that are authorized to make final judgments on how existing laws apply to specific individual cases, and that there is no guarantee that the views described in this Guidance will be adopted by the courts without change.

1.3 Definitions of terms

The definitions of primary AI-related terms used in this Guidance are as follows, all of which are substantially the same as the definitions found in the AI Guidelines for Business.[11]

Term	Definition
AI	As it is difficult to strictly define AI, we use the same definition as in the AI Guideline for Business, as follows: An abstract concept, which includes AI systems (see below) themselves or software or programs that perform machine learning.
AI service	As in the AI Guidelines for Business, “AI service” means a service that uses AI systems. It refers to providing AI business users with values in general. AI services are provided and operated through not only technologies that constitute an AI system, but also non-technological approaches, including human monitoring and appropriate communications with stakeholders.
AI system	As in the AI Guidelines for Business, “AI system” means a system (such as a machine, robot, and cloud system) that works at various levels of autonomy during the use process and incorporates a software element that has a learning function.
AI model	As in the AI Guidelines for Business, “AI model” means a model incorporated into an AI system and acquired through machine learning using training data. It produces prediction results in accordance with the input data.
AI developer	As in the AI Guidelines for Business, “AI developers” means business operators who develop AI systems (including business operators who research AI). They develop AI models as well as

[11] AI Guidelines for Business, main body pp.5 and 9-10.

	algorithms and contribute to construction of AI systems including AI models, base system, as well as I/O functions via data collection (including purchase), data preprocessing, training with data. In addition, even after the development and operational deployment of AI models and systems, they are also responsible for maintaining and improving model performance through post-training (post-training), which aims to expand domain knowledge in specific areas, adapt to changes in the environment, and further make adjustments (alignment) to ensure behaviors aligned with human intentions and values.
AI provider	As in the AI Guidelines for Business, “AI providers” means business operators who incorporate AI systems into applications, products, or existing systems, business processes, etc., and provide them to AI business users and, in some cases, non-business users as services. They verify AI systems, integrate AI systems with other systems, provide AI systems and services, offer operation support for AI business users on AI systems for normal operations, or perform the AI service operation itself.
AI business user	As in the AI Guidelines for Business, “AI business users” means business operators who use AI systems or AI services in their businesses. Their role is to use an AI system or AI service in an appropriate way intended by the AI provider, share information such as environmental changes with the AI provider, continue the normal operation, operate the provided AI system as necessary. In addition, when non-business users might be affected by AI use in some ways, AI business users are also responsible for making efforts to prevent AI from incurring unexpected disadvantages for those non-business users and maximize benefits from AI.

Chapter 2 General Principles

2.1 Subjects of discussions by the Study Group

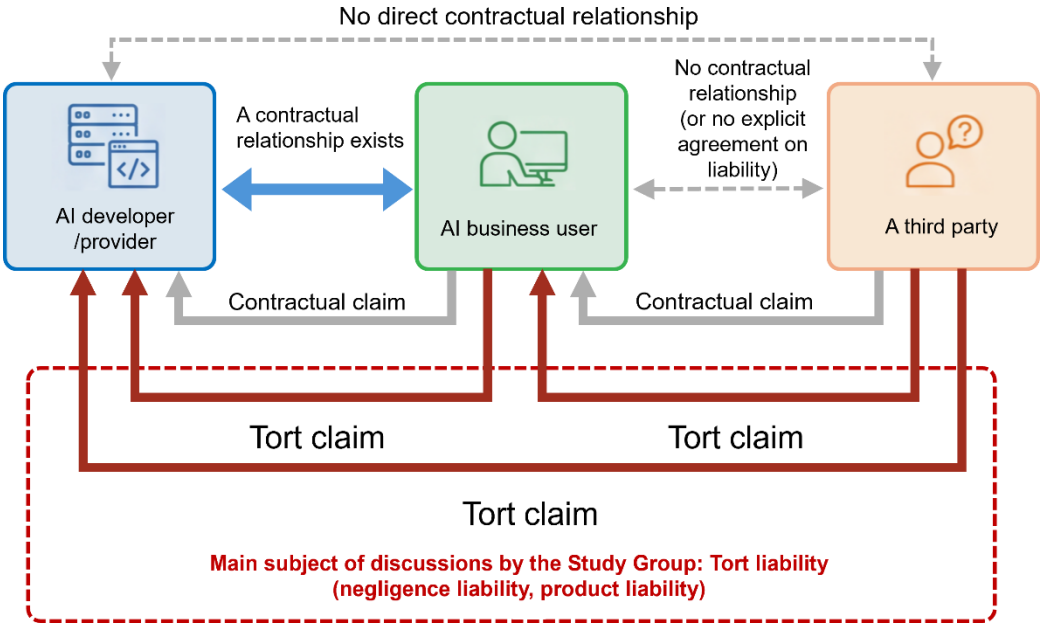
The legal grounds for liability for any damage caused by the utilization of AI mainly include tort liability, which arises without a contract, and contractual liability, which arises under a contract between the relevant parties. Of these grounds for liability, tort liability was the focus of discussions by the Study Group.[12] The reasons for this are as follows:

(A) Tort liability is a default rule for liability between the parties regardless of whether or not a contract exists. In particular, it provides a primary remedy for the rights of a damaged third party who is not a party to a contract.
(B) Parties to a contract can agree to increase, limit, waive, or otherwise change tort liability under the contract. Presenting certain views on tort liability may be useful in distributing liability or loss among parties under a contract.
(C) Details of contractual liability depend on the provisions of a contract between the parties and are highly specific. The Contract Guidelines and the Contract Checklist have presented certain views on contractual considerations relating to the development, provision, and utilization of AI.
(D) While tort law does require some interpretation and application on a case-by-case basis, the rules and value judgments at issue are often common in each category of cases.

The schematic diagram below shows an overall picture of the legal relationships between the parties involved in the utilization of AI, including contractual liability. The hypothetical cases used in Chapters 3 and 4 discuss legal relationships that are substantially in line with the diagram below.[13]

[12] In some cases, the Study Group also discussed due care of a prudent manager or duty of care under a contractual relationship between an AI business user and a third party, which is often an issue together with tort liability.

[13] Various patterns are possible in reality, such as cases where the AI developer and the AI provider are separate entities or where the AI provider and the AI business user are the same entity. These cases are omitted in the schematic diagram.



The subjects covered by this Guidance are limited to cases where damage is incurred by the third party shown in the above diagram. In these cases, the third party often has no contractual relationship with the parties involved in the development, provision, and utilization of AI or, if it does, the contract often contains no explicit provisions on who is liable for the damage. Therefore, making claims under tort law is an important option.

On the other hand, there can be cases where damage is incurred by the AI business user due to a failure of AI to perform as agreed between the AI developer/provider and the AI business user. The legal relationship in these cases depends on the terms of the contract between the AI developer/provider and the AI business user. As this requires an analysis based on the terms of the specific contract, these cases were basically not a subject of discussions by the Study Group.[14]

While various legal provisions provide grounds for tort liability, this Guidance mainly discusses general tort (Article 709 of the Civil Code) and product liability (Article 3 of the Product Liability Act), which apply to a wide variety of cases and are particularly likely to be issues in the context of the utilization of AI.

2.1.1 General tort

The requirements for general tort liability under the Civil Code (Article 709 of the Civil Code) have been organized in different ways. The following is one way to organize the requirements.

[14] However, in Case C discussed in hypothetical case 6 in 4.2 below, the Study Group discussed product liability of the manufacturer to the business using a robot on the assumption that there is no explicit agreement between the parties.

Requirement	Description
(A) Negligence	The actor violated a normative duty to avoid the occurrence of damage (duty to avoid the consequences) even though the actor could have foreseen the occurrence of the damage (i.e., the foreseeability of the damage).
(B) Infringement of protected legal interests	Bodily integrity, property, moral rights, or other interests that deserve legal protection are found to have been infringed.
(C) Occurrence of damage	The victim has incurred damage.
(D) Causal relationship	There are an actual causation (i.e., a "but-for" relationship) and a legal causation (i.e., a relationship in which the act usually results in the consequences; this requirement is often included when applying Article 416 of the Civil Code by analogy) between requirement (A) and requirement (B) above.

With regard to requirements (B) through (D) of the above requirements, the conventional concepts are still often valid in settings where AI is utilized.[15] On the other hand, it is often difficult to decide how negligence (requirement (A)) should be understood during the use of AI. As described above, negligence is determined based on whether or not the actor is found to have violated a duty to avoid the consequences subject to the foreseeability of the damage. The level of this duty of care is determined based on, among other things, the size of the risk arising from the tort, the gravity of the infringed interests, and the attention of a reasonable person who is in the occupation or position.[16]

The conventional concept of negligence assumed that a tort is committed by a person, and the focus was whether or not that person is found to have violated a duty of care in making-decisions or engaging in acts. However, in the context of the autonomy of AI (i.e., the ability to operate without human intervention to a certain extent), which is one of the important characteristics of AI, how the negligence of AI developer, provider, and business user should be understood becomes a significant issue, particularly in settings where no one is making decisions or engaging in acts.

In addition, in cases where multiple parties contribute to the occurrence of damage, as opposed the above cases of a tort by a single person, whether or not a joint tort (the first sentence of paragraph (1) of Article 719 of the Civil Code) is committed or whether or not liability for aiding

[15] In the context of AI, it could also be difficult to find a causal relationship between negligence and rights infringement. See 5.1.4 below.

[16] Kato, Ichiro (ed.) (1965). *Chushaku minpo [Japanese civil law annotated] (19), Saiken [Claims] (10) (in Japanese)*. pp. 24-25 [Kato, Ichiro]. Yuhikaku.

and abetting (paragraph (2) of the same article)[17] arises becomes an issue.[18] As with general tort liability, there are different ways of organizing the requirements for joint tort liability, and the following is one way to organize the requirements. If the following requirements are met, the parties are liable for damage jointly and severally with each other.

Type	Requirement	Description
Joint tort (in a narrow sense)	(A) A tort committed by two or more tortfeasors	The tort committed by each of the tortfeasors meets the requirements for a tort (see the above table).
	(B) A joint nature	A piece of damage was caused jointly by the acts of two or more tortfeasors.[19]
Aiding and abetting	(A) A tort by another person	A tort committed by another person meets the requirements for a tort (see the above table).
	(B) The act of aiding and abetting	Providing physical and/or mental assistance that facilitates the tort (aiding and abetting).

2.1.2 Product liability

Product liability law is a special law of the tort law under the Civil Code described in 2.1.1 above. The purpose of product liability law is to reduce the victim's burden of proof by acknowledging the manufacturer's liability for damage regardless of whether or not there was negligence, in cases where defects in the manufacturer's product caused damage to the life, body, or property of others. As with tort liability, the requirements for product liability have been organized in different ways. The following is one way to organize the requirements.

---

[17] While there are not many judicial precedents in which liability for aiding and abetting was acknowledged, the following are examples in which the court acknowledged the liability of a person(s) who provided important tools and services to the direct tortfeasor.

[1] Supreme Court Judgment, March 2, 2001, *Minshu (Supreme Court Civil Case Reports) Vol. 55, No. 2*, p.185 (a case in which the court acknowledged the tort liability of a karaoke equipment lease company that failed to check that the lessee had signed or applied for a copyrighted work license agreement; reference judicial precedent 6 below).

[2] Tokyo District Court Judgment, March 23, 2016, *Hanrei Jiho, No. 2318*, p.40 (a case in which the court ordered compensation for damage from the persons who sent their identity documents or the like to their employer who requested them under the pretext of necessity for their "side job," on the grounds that they could have easily become aware that their activities were likely to be being used for some illegal activities, even though they were not aware that they were aiding and abetting the opening of accounts for fraudulent business practices).

[3] Tokyo High Court Judgment, December 20, 2017, *Hanrei Jiho, No. 2384*, p. 20 (a case in which the court acknowledged that the person who leased an office to a fraudulent investment company committed aiding and abetting by negligence).

[18] This discussion does not cover (i) joint torts in which the tortfeasor is unknown as set forth in the second sentence of paragraph (1) of Article 719 of the Civil Code or (ii) abetting as referred to in paragraph (2) of the same article, as these are unrelated to the hypothetical cases discussed by the Study Group.

[19] According to judicial precedents, it suffices if the tortfeasors objectively acted jointly with each other and it is not a requirement that they subjectively acted jointly with each other, such as in conspiracy (see, for instance, Otsuka, Tadashi (ed.) (2022). *Shin Chushaku Minpo [Japanese civil law annotated, new edition] (16) Saiken [Claims] (9) (in Japanese).* p. 290 [Otsuka, Tadashi]. Yuhikaku; and Yoshimura, Ryoichi (2024). *Fufo Koui Ho (Dai 6-pan) [Tort law (sixth edition)] (in Japanese).* p. 262. Yuhikaku).

Requirement	Description
(A) Manufacturer or the like	The entity that receives a claim is a person who manufactured, processed, or imported the product in the course of business or who falls under any of the items of paragraph (3) of Article 2 of the Product Liability Act.
(B) Product	The product in question is a manufactured or processed movable.[20]
(C) Delivery	The manufacturer or the like voluntarily transferred its possession of the product.
(D) Defects	The product lacks the safety that it should normally have.[21]
(E) Damage to life, body, or property	The life, body, or property of others was damaged.
(F) Causal relationship	There are an actual causation (i.e., a “but-for” relationship) and a legal causation (i.e., a relationship in which the act generally results in the consequences; this requirement is often included when applying Article 416 of the Civil Code by analogy) between requirement (D) and requirement (E) above.

Under Japanese product liability law, product liability only applies to movables (requirement (B) of the above table). Therefore, even though product liability does not apply to AI or other software, product liability may arise from a product incorporating software. In the future, machines, robots, etc. that are autonomously operated by software, including AI, are expected to become widespread. An important issue under these circumstances is how to determine whether defects (i.e., a lack of the safety that the product should normally have; requirement (D) of the above table) exist, which is the primary requirement, in light of the characteristics of the relevant product.

2.1.3 AI governance and the AI Guidelines for Business

When determining liability as described above, an important consideration is the concept of AI governance for risk control that takes the autonomy, uncertainty, etc. of AI into account. In Japan, the AI Guidelines for Business have been established and published regarding the details of AI governance required of businesses engaging in the development, utilization, and provision of AI. The Guidelines are particularly useful in determining general tort liability described in 2.1.1 above.[22]

[20] Article 2, paragraph (1) of the Product liability Act.

[21] Article 2, paragraph (2) of the Product liability Act.

[22] Since the AI Guidelines for Business cover the governance of AI in general, the guidelines can be useful in terms of products incorporating AI, to which product liability applies. However, at the time of writing this Guidance, the AI Guidelines for Business do not discuss these products in detail. However, in this regard, the risks involved in physical AI and the details of governance that should be conducted by each relevant party are discussed, among other things, on p. 17 et seq. of “AI Jigyosha Gaidorain no Koshin ni Muketa Ronten [Points at issue toward updating the AI Guidelines for Business]” which is in Shiryo 4 [Handout 4] for the 28th meeting of the Study Group for AI Governance under the Conference toward AI Network Society of the Ministry of Internal Affairs and Communications (December 2, 2025), as part of the discussions for updating the AI Guidelines for Business in light of the trends in physical AI, which integrates AI in the form of software (AI algorithms) and physical hardware (sensors, actuators, edge devices, etc.) (https://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02tsushin06_04000132.html; last accessed on January 6, 2026).

The AI Guidelines for Business provide a summary of the basic concept of AI governance required of businesses in line with the actual situation in Japan, while also taking into account the trends in international AI governance. The main body of the Guidelines lists the common principles, such as safety, fairness, and transparency, that each business operator should take into account when setting targets, and requires business operators to develop AI governance based on a risk-based approach that serves to achieve the targets. Specific methods for developing AI governance are presented in the Appendix to the Guidelines. When developing AI governance, it may be difficult to foresee and control risks in advance due to the rapid progress of AI technology. For this reason, the AI Guidelines for Business takes an agile governance approach, in which: potential risks are organized based on the environment in which AI is used and on the uses of AI, before conducting environmental and risk analyses; based on the results of the analyses, an AI management system is designed; and the management system is continuously assessed through appropriate feedback on risks and troubles emerging in the process of operation of the system.

The AI Guidelines for Business intends to encourage business operators to make voluntary efforts. Due to this nature, the Guidelines do not require compliance with everything contained in them nor do they provide exhaustive descriptions of the measures required for each setting. Therefore, conforming or failing to conform to the AI Guidelines for Business does not immediately mean there was or was not negligence. However, in light of the framework for determining negligence as described in 2.1.1 above, if a business operator has examined and analyzed risks based on, among other things, the approach taken by the AI Guidelines for Business and has created a system for conducting such examination and analysis, and if any peculiar risk that was difficult to foresee still emerges, then the business operator's efforts in line with the Guidelines would be taken into account as circumstances that work to deny the foreseeability of the risk. Similarly, if a business operator has taken practically reasonable measures to deal with expected risks according to their levels, these efforts would be taken into account as circumstances that reduce the possibility of the business operator being considered to have violated a duty to avoid the consequences.[23] Specific cases in which this approach is appropriate are described in 3.3, 3.4, 4.1, etc. below.

[23] As part of AI governance, utilization of certifications based on international standards for AI management systems may also serve as a possible measure. For example, in an attempt of an AI user to establish an appropriate framework for the use of AI, if the user confirms whether the AI developer or provider has obtained certifications related to AI management systems such as ISO/IEC 42001, obtains the developer's or provider's evaluation regarding the performance, limitations, and risks of the AI products, AI services, or AI systems to be introduced, and verifies their suitability for the company's operational requirements and the appropriateness of their use environment, such efforts may be taken into account as one of the circumstances showing that the user implemented practically reasonable measures. However, as with the establishment of AI governance based on the AI Guidelines for Business, such measures do not automatically determine the presence or absence of negligence.

2.2 Assistive/supportive AI and dependable/substitutive AI

Based on the Study Group's discussions on the hypothetical cases, it is useful to discuss how to determine liability of each of the parties involved by dividing AI into two categories according to the manner in which it is used.

The first category, assistive/supportive AI, is used exclusively to assist or support AI business users' decision-making and is intended to eventually involve human judgment or action. Currently, most AIs seem to fall under this category. Assistive/supportive AI is able to evaluate whether or not an AI business user used the AI by exercising the due care that should properly be exercised by an AI business user, by assessing the appropriateness of the final human judgment or action. This allows this category of AI to apply the conventional framework for determining negligence (2.1.1 above).

On the other hand, in light of the recent progress of AI technology, another category of AI, or dependable/substitutive AI, seems to be appearing. This category of AI does not necessarily require the involvement of the AI business user's final judgment or action, but (i) is provided on the assumption that it will substitute all or part of human judgment or action; and (ii) is expected to be used by relying on the AI's judgment. Since this category does not necessarily involve the AI business user's judgment or action in the context of the rights infringement or damage that eventually occurred, it has a visible problem in that it comes with difficulty in determining negligence by assessing the AI business user's judgment or action.

In addition, the level of duty of care expected of AI developers or AI providers who sell or provide AI systems or AI services would be different for different categories as described below.

2.2.1 Assistive/supportive AI

(1) AIs falling under the category of assistive/supportive AI

The category of assistive/supportive AI is used exclusively to assist or support AI business users' decision-making and is intended to eventually involve human judgment or action. Due to their nature or manner of use, some AIs are considered not suitable for substituting human judgment or action but meant to be used as assistive/supportive AI. Whether or not this is the case with a given AI depends on the AI's technological characteristics and their trends, applicable regulatory laws, and other factors. As of the time of writing this Guidance, possible cases include: (A) cases where the AI cannot be regarded as substituting human judgment in light of the AI's functions and the settings where it is used; (B) cases where the final human judgment is required for reasons specified by regulatory law; and (C) cases where the AI's output entails a potential risk of infringing third-party rights, and this requires human assessment or examination. Specific examples for each category of the above cases are described in Chapter 3 below.

AIs that do not fall under any of (A) through (C) above may be both assistive/supportive AI and dependable/substitutive AI. However, AIs that are not dependable/substitutive AI based on the criteria described in 2.2.2 below are generally assistive/supportive AI.

(2) Liability of AI business users using assistive/supportive AI

As described above, since assistive/supportive AI eventually involves human judgment or action, it is able to apply the conventional framework for determining negligence (2.1.1 above) by assessing the human judgment or action. This means that when an AI business user uses assistive/supportive AI, it is always used to assist or support the user's decision-making, and the user's liability is determined based on whether or not the user made an appropriate judgment or took an appropriate action by exercising the due care that should properly be exercised by an AI business user in the relevant setting, regardless of whether or not AI was used. The fact that the AI business user adopted the AI's output does not, in principle, increase or decrease the level of duty of care required of the AI business user.[24]

Under the general duty of care described above, AI business users are required to use the AI's output to an appropriate extent in making the final judgment or taking the final action. However, factors that are important in determining the user's negligence may vary according to the characteristics of the AI or the nature of the risks involved. Typically, AI business users are required to use the AI by assessing the correctness and appropriateness of the AI's output when taking the output into consideration (hypothetical case 1 in 3.1 and hypothetical case 2 in 3.2 below). However, there are use cases where the AI's output does not at a glance clearly indicate where risks lie and it is not always easy for the AI business user to examine and correct the output (hypothetical case 3 in 3.3 below and hypothetical case 4 in 3.4 below). In the latter cases, the AI business user's duty of care may include collecting information to a certain extent in advance and taking advance measures necessary for the user's use, in order to prevent third parties from suffering rights infringement or damage as a result of the user's eventual use of the AI's judgment. As for AI developers/providers, it becomes an issue whether or not they should have taken certain design measures (see (3) below).

(3) Liability of AI developers/providers who sell or provide assistive/supportive AI

Assistive/supportive AI assists or supports people in making the final judgment or taking the final action. Due to this nature, this category of AI is considered an information processing system that provides a certain output based on the results of analysis of the input data and prompts. In this type of system, even if the AI provides an inappropriate output, the AI business user is basically supposed to eventually examine the appropriateness of, and correct, the AI's output.

[24] However, the opinion was expressed at a meeting of the Study Group that it is important to discuss the view that using an AI system of particularly high accuracy may have an impact on the nature of the AI business users' duty of care (see 3.2.2 below).

Therefore, there are only limited settings where AI developers/providers who sell or provide assistive/supportive AIs are held liable to third parties.

In the conventional argument on tort law, the possible causal relationship was considered to be denied in principle if another person's judgment was involved in the rights infringement, on the grounds that the other person's decision was the direct cause.[25] Reference judicial precedent 1 described below provides an example where the manufacturer's liability was denied in relation to a car navigation system, which is one example of the conventional information processing system, on the grounds that the driver, who is the receiver of information, should eventually determine safety in an independent manner.

Reference judicial precedent 1: Fukushima District Court Judgment, December 4, 2018, *Hanrei Jiho, No. 2411,* p. 78.

- **In this case, the plaintiff was driving according to the route guidance from the car navigation system manufactured by the defendants, when the car entered a narrow road with trees growing into the road and suffered scratches. The plaintiff claimed damage from the defendants on the grounds of tort and product liability.**
- **For the following reasons, the plaintiff should have driven along the road in question on the plaintiff's own judgment instead of relying on the route guidance from the car navigation system: (A) it is impossible or extremely difficult for a car navigation system to provide real-time information on the accurate condition of roads across the country; (B) the driver who actually faces individual roads is best able to monitor their safety; (C) the screen and user's manual of the car navigation system show warnings and the like that the driver should drive according to the actual conditions on the road and other circumstances; and (D) the road where the accident occurred is described as a "light vehicle road" on the topographic map published by the Geospatial Information Authority of Japan, which means that the car navigation system did not contain information on a road that vehicles cannot pass through. Therefore, the causal relationship between the route guidance and the scratches was denied, and the defendants were not found liable.**

As seen in the above argument and judicial precedent, the potential causal relationship between the system's output and the rights infringement or damage is denied in principle if the system is intended to eventually involve human judgment or action. In addition, negligence would not be acknowledged if the system was provided after taking appropriate explanatory and design measures from the viewpoints described below.

In settings where an AI is used, it is important that AI business users appropriately understands the AI's performance limits and other characteristics and is able to use it without causing excessive risks. In this regard, some judicial precedents acknowledge the liability of the developer or provider of a product used by others if the product caused damage during the others' use of the product. Specifically, discussions have been undertaken as to: (A) liability in cases where the

[25] Kubota, Atsumi (ed.) (2024). *Shin Chushaku Minpo [Japanese civil law annotated, new edition] (15) Saiken [Claims] (8) (in Japanese).* pp. 387-388 [Hashimoto, Yoshiyuki]. Yuhikaku.

user's misuse was caused by a failure to explain information that is important for the user's appropriate use of the product;[26] and (B) liability in cases where the product was provided without appropriate measures to prevent rights infringement, despite a high risk of rights infringement anticipated during use of the product.[27] From these viewpoints, there is room for discussing the liability of AI providers or AI developers.

The duty to explain described in (A) above would be relatively more important with assistive/supportive AI. The performance limits of an AI and how to use it are not self-evident to AI business users. If proper explanations are not given, the functional or performance limits of the AI may not be appropriately understood by AI business users, in which case they may develop excessive trust in the AI. In judicial precedents, the courts identify the duty to explain after discussions that take into account the nature and level of the risks involved, the level of uneven distribution of information between the parties, and the attributes of the user, among other factors. It would be important for AI developers/providers to provide clear explanations about the functional and performance limits of their AI, how to use it, and important risks involved, etc., to the extent reasonably foreseeable in light of the state of scientific and technical knowledge at the time, by taking those factors into account (the duty of care in providing explanations).[28] As long as such explanations have been provided, the AI business user is expected, in the context of division of roles between the parties, to examine and correct any inappropriate output of the AI, and negligence is not acknowledged with respect to the AI developer and the AI provider in principle.

The liability described in (B) above basically becomes an issue only when it is not always easy for the AI business user to foresee specific risks or to use the AI by controlling its output, considering the functions and nature of the AI, the characteristics of the AI business user, etc. In such case, it becomes an issue whether the AI developer/provider should have taken certain measures to prevent rights infringement or such other action (the duty of care in design; for specific examples, see 3.3.4 and 3.4.4 below).

---

[26] For instance, a case involving a lack of explanations on important information on a medical device was discussed in the Tokyo High Court judgment dated February 7, 2002, *Hanrei Times, No. 1136*, p. 208. The point at issue in this judicial precedent was the manufacturer's liability relating to the blood pump in a heart-lung machine. The judgment was passed before the enforcement of the Product Liability Act. The court acknowledged the manufacturer's violation of its duty to explain under general tort liability.

[27] For instance, a case involving a lease of karaoke equipment in a situation where there was a high probability for copyright infringement was discussed in the Supreme Court judgment dated March 2, 2001, *Minshu (Supreme Court Civil Case Reports) Vol. 55, No. 2,* p. 185 (reference judicial precedent 6 described below).

[28] The opinion was expressed at a meeting of the Study Group that, in light of the fact that an AI is basically defined by three factors—training data, prompts input by the user, and external databases used by RAG or the like—it is important to explain the following for the purpose of reducing the AI service provider's risk of violating its duty to explain: what data is used to train the AI; the scope of judgments that can be made by the AI by using which database; and the limitations of the AI. Explanations by the AI developer or the AI provider do not have to always cover all the factors; rather, reasonable explanations should be provided within a reasonable scope, taking into account the general considerations regarding the duty to explain described in the main text (such as the nature and degree of risk, the degree of information asymmetry between the parties, and the attributes of the users). Although appropriateness of such explanations would have to be ultimately made on a case-by-case basis, one potential approach is to judge whether or not the explanations were appropriate in light of the general knowledge in the industry.

The content of the duty of care in design explanation required of AI developers and AI providers may vary depending on their respective roles within the value chain.

2.2.2 Dependable/substitutive AI

As described above, it is difficult to apply the conventional approach to the category of dependable/substitutive AI without change. The issue is whether an approach different from that for assistive/supportive AI can be taken to determine liability for dependable/substitutive AI and, if so, to what cases the different approach should be applied.

(1) Approach to determine liability for AIs falling under the category of dependable/substitutive AI

(A) Judicial precedents involving a system that is not intended to involve human operation or intervention

While there have long been various machines and systems that are intended to increase the efficiency of human work, most of them intended to involve human operation or intervention, particularly in settings where third-party rights or interests may be affected. On the other hand, the progress of AI technology has brought the ongoing advent of systems that provide high performance and safety without human operation or intervention, even in areas where third-party rights or interests are involved.

With respect to the issue of how the AI business user's negligence should be determined in such case, there are no judicial precedents that directly dealt with this issue and there is no established approach, which makes this issue difficult. The following judicial precedent involves a system that does not intend to involve human operation or intervention in individual processing.

Reference judicial precedent 2: ATM withdrawals by unauthorized persons (1) (Supreme Court Judgment, July 19, 1993, *Shumin (Civil Chapter of the Cases of the Supreme Court) No. 169,* p. 255)

- **If an unauthorized person withdrew money from an ATM by using a passbook or bank card and entering the PIN, and if the ATM was a bank ATM, the bank card was genuine, and the correct PIN was entered, then the bank's disclaimer in the terms and conditions is valid, unless there were exceptional circumstances on the part of the bank, such as poor management of the PIN.**
- **The magnetic stripe on a bank card that was used at that time contained the encoded registered PIN, which could be decoded using a commercially available card reader or the like. However, considering that reasonable knowledge and skills in computers are necessary to decode the PIN in such manner, the bank card does not lack security to such an extent that the validity of the disclaimer must be denied.**

Reference judicial precedent 3: ATM withdrawals by unauthorized persons (2) (Supreme Court Judgment, April 8, 2003, *Minshu (Supreme Court Civil Case Reports) Vol. 57, No. 4,* p. 337)

- **In order for the bank to be free of negligence in the event of an unauthorized withdrawal, the bank must have fulfilled its duty of care in order to eliminate withdrawals by unauthorized persons to the extent possible, in the bank's overall management of installation of the money withdrawal system using ATMs, including expressly stating to depositors that ATM withdrawals are available.**

While the judicial precedents described above do not involve tort law, they present rulings on the requirement of being "without negligence" as referred to in Article 478 of the Civil Code, which are useful in determining negligence under tort law.

The approach taken by the Supreme Court across reference judicial precedents 2 and 3 can be organized as follows. In the case of a money withdrawal system, if the genuine passbook or bank card and the PIN for it fall into the hands of another person, there is a risk that an unauthorized person may make withdrawals. However, if this is so unusual as to be exceptional, the bank is not immediately held negligent for any mistaken payment[29] (reference judicial precedent 2). Then, in order for the bank to be free from negligence, the bank must have fulfilled its duty of care in order to eliminate withdrawals by unauthorized persons to the extent possible, in the bank's overall management of installation of the money withdrawal system (reference judicial precedent 3). This framework for determination of negligence has also been used in judicial precedents involving an Internet banking or similar system operated by a bank.[30]

The approach taken by the judicial precedents described above can be understood as follows: assuming that the money withdrawal system, which carries out payment and other processing without human involvement, has a certain level of security, negligence should be determined by looking at the process of installation and operation of the system, instead of at human judgment or action.

Dependable/substitutive AI is no different from the money withdrawal system which processes tasks without final human judgment or action. Therefore, the framework for determination of negligence described above would also apply, mutatis mutandis, to the determination of liability in cases where dependable/substitutive AI is used. Specifically, in the case of dependable/substitutive AI, negligence is determined based on whether the AI business user has appropriately created the work process incorporating an AI system, and whether the AI business user performs reasonably possible operation to reduce the possibility of the AI's undesirable output causing rights infringement or damage, unlike in the case of assistive/supportive AI where the appropriateness of final human judgment or action is at issue.

---

[29] What is directly presented by reference judicial precedent 2 is a ruling on the validity of a special disclaimer clause between the bank and the depositor, but since this type of special disclaimer clause is considered to be a materialization of the no-fault requirement under Article 478 of the Civil Code, this precedent can be considered to substantially judge whether or not the bank was at fault (Yoshida, Mitsuhiro (2003), "Hanrei Hihyo [Case comment]," *Hanrei Times, No. 1124,* p. 88).

[30] For instance, the Tokyo High Court judgment dated July 13, 2006 (*Kinyu Homu Jijyo [Banking Law Journal], No. 1785,* p. 45) supported the application of the disclaimer to mistaken bank transfers that occurred through Internet banking, on the following grounds: if the PIN and other details matched at the time of money transfer, the bank is not held liable unless exceptional circumstances existed, such as the bank's poor PIN management, whereas the bank had created and managed the system such that the system as a whole can eliminate transfers by unauthorized persons to the extent possible.

(B) Requirements for AIs to fall under the category of dependable/substitutive AI

Reference judicial precedents 2 and 3 assumed that the money withdrawal system was reliable to some extent for the purpose of processing payment and receipt of the depositor's assets. When applying mutatis mutandis the approach to determine liability described above to dependable/substitutive AI, the important point is what AI would make it reasonable to leave judgment to the AI. It is expected that this will be the point at issue in actual cases.

As described in 2.2.1(1) above, some AIs should be used as assistive/supportive AI due to their nature. Other AIs may be both assistive/supportive AI and dependable/substitutive AI. Whether an AI falls under the category of dependable/substitutive AI would be determined on a case-by-case basis based on the following two factors.

(a) Necessity

First, in order for an AI to fall under the category of dependable/substitutive AI, there must be a necessity to leave judgment to the AI without the involvement of final human judgment or action. Specifically, it is required that the AI be expected to bring benefits that are difficult to achieve if human judgment or action is involved by the AI business user, such as cases where the AI processes high loads of tasks that are difficult for people to do entirely, as with the money withdrawal system described in (1) above, or where the AI automatically processes tasks in scenes where it is temporally or physically difficult for people alone to do the tasks. When it is not necessary to leave judgment or action to AI, people are able to use AI by examining and correcting its output by exercising necessary care, in which case people may have a duty to avoid consequences by having human judgment or action involved.

(b) Accuracy and safety

The next important issue is what level of accuracy and safety dependable/substitutive AI should have when we leave human judgment or action to this category of AI. Abstractly speaking, dependable/substitutive AI would have to meet socially expected levels of accuracy and safety. For instance, an AI-based system that meets applicable regulatory law or applicable safety standards in the industry, or that has been introduced and is commonly used across society, is more likely to be assessed as having the level of accuracy expected of dependable/substitutive AI.

In the current situation where AI-based systems are being spread, there are no clear standards for the level of accuracy or safety that is sufficient for them to fall under the category of dependable/substitutive AI. However, there would already be some AI systems with sufficient accuracy and safety at this point in time. Under tort law, efforts have been made to fairly distribute damage in settings where rights infringement or damage occurred, by: (i) setting the standards for the judgment and action of the person who caused the damage at the level of care that would be exercised by a reasonable person who is in the situation or position in which the damaging person was placed, in light of, among other factors, the size of the risk involved in the damaging person's activities; and (ii) finding that the damaging person was negligent if the person's judgment or

action is below the standards (2.1.1 above). This suggests that when introducing AI to replace human activities, we can set the standard at the level of work by a reasonable person engaging in the same activities. If an AI's work level is assessed as equivalent to or higher than the conventional level of work by a reasonable person in terms of quality and accuracy, or if the risk of rights infringement is adequately controlled as compared to the conventional level of work by a reasonable person, it would be found reasonable to use the AI to replace all or part of human judgment or action (or to determine that the AI falls under the category of dependable/substitutive AI).[31, 32]

As a practical method to substantiate the applicability to dependent/alternative AI, one may consider, as appropriate, confirming that the AI meets the standards of work quality and safety historically required for the tasks in which it is to be deployed, and conducting additional verification in the actual operational environment — in each case taking into account benchmark evaluations provided by the AI developer or provider, known performance limitations and erroneous output patterns, discrepancies between the test environment and actual operational conditions, and other relevant factors.

It is also expected that there may be cases where an AI system works with a level of accuracy that is equivalent to or higher than a reasonable person only in part of the work process or where an AI system works with a reduced level of accuracy that is lower than the level of work by a reasonable person under certain conditions. In these cases, operating the AI system by replacing human judgment with AI to a certain extent can be found reasonable if the overall process operates at a work level equivalent to or higher than human beings by combining the AI with human involvement, the work process, the existing IT infrastructure, etc., such as, for instance, by assessing and correcting the AI's output by human monitoring or using another system to a certain extent, or by creating an environment for the AI to work with the intended level of accuracy. With the future progress of technology, the number of settings may increase where an AI system is able to universally operate the work process. However, such use cases are limited at this point, and systems that partially substitute human judgment or action as described above or that

[31] It should be noted that while the conventional level of work by a reasonable person as described in the main text will serve as a comparison standard for the time being in areas where dependable/substitutive AI has scarcely been introduced or is being spread, once using systems with higher accuracy becomes socially expected with the spread of AI and other autonomously operated systems, systems of the same types with high reliability may serve as comparison standards.

[32] For the purpose of protecting the rights of third parties, the work level for AI as described in the main text would be the primary standard for assessment. However, in some cases, it is difficult to simply compare the accuracy of AI and the level of work by a reasonable person. In such cases, we could weigh the potential benefits, such as an increase in work efficiency, brought by an AI against the potential risks involved in the AI (i.e., the probability of occurrence and the level of severity of rights infringement and damage caused by the AI) and, if the former exceeds the latter, take the potential benefits into account as an additional factor that supports the use of the AI as dependable/substitutive AI, except in cases where serious infringement of legal interests, such as damage to life or health, might be brought by the use of the AI. For useful discussions, see, for instance, the view on socially useful activities associated with abstract dangers, to the effect that consideration based on the principle of proportionality should be given to the duty to act (i.e., the duty to foresee and the duty to collect information) in the context of negligence in the stage of endangerment. Otsuka, Tadashi et al. (2023) *Minpo 6: Jimu Kanri, Futo Ritoku, Fuho Koui [Civil law 6: management without mandate, unjust enrichment, and tort] (in Japanese)*. p. 140. Another possibility would be to take into account an AI's satisfaction of the requirements under relevant regulatory law as an additional factor that supports the position that the AI provides the required safety.

autonomously operate in certain environments are more likely to be assessed as meeting the required standards. Specific examples are mentioned in 4.1 and 4.2 below.

If an AI meets the two requirements described above, and if the AI is provided with the intention of substituting human judgment or action, then we believe that negligence should be determined assuming that the AI falls under the category of dependable/substitutive AI.

(2) Liability of AI business users using dependable/substitutive AI

When the application of the approach taken with respect to the ATM system described in (1) above is extended, liability of AI business users who use dependable/substitutive AI is transformed from the duty of care to make reasonable human judgment and take reasonable human action into the duty to properly build a work process incorporating the AI system and to operate it with as low risk as possible.

(A) Duty of care to build a work process incorporating an AI system

When introducing dependable/substitutive AI, one has to consider replacing part of the human judgment in the work process, the existing IT system, etc. with the AI system. An AI business user's duty of care in this context is considered from aspects such as whether the relevant work process is suited for automation and whether the AI system used meets the desired standards and provides desired safety.[33, 34]

Another important factor is whether or not an appropriate system for using AI and appropriate governance have been developed in light of the risks involved in AI. For instance, in cases where an AI system works with a level of accuracy that is equivalent to or higher than a reasonable person only in part of the work process or where an AI system works with a reduced level of accuracy that is lower than the level of work by a reasonable person under certain conditions as described in (1) above, the AI business user may be required to ensure that the overall work process operates with desirable accuracy and safety, by introducing human involvement and/or integrating the system with another system. In such case, the AI business user would have a duty of care to consider human involvement and/or another system necessary to achieve the desired accuracy and safety and to introduce necessary human and/or physical equipment.

(B) Duty of care in operating the work process so built

Assuming that the work process has been built as described in (A) above, another factor to consider when determining negligence of an AI business user is whether or not the AI business

---

[33] Since an AI may change in accuracy and other metrics afterward through continued training and updates, another type of duty of care may arise as to whether an AI is no longer fit for use as dependable/substitutive AI due to reduction in accuracy through the process of operation.

[34] In cases where there are two or more choices of AI candidates for introduction once dependable/substitutive AI has become more widespread as described in note 31 above, another type of duty of care may arise as to whether the AI business user uses an AI with the more desirable performance.

user has taken reasonable measures in the operation of the system in order to reduce the possibility of rights infringement caused by undesirable AI output. For instance, since an AI may change in accuracy and other metrics afterward through continued training and updates, the AI business user could, among other things, check whether the AI is no longer fit for use as dependable/substitutive AI due to reduction in accuracy through the process of operation.

Specific measures to be taken with respect to the types of duty of care described in (A) and (B) above are different for different AI systems. The AI Guidelines for Business requires governance to control the risks involved in the utilization of AI. AI business users could take reasonably possible measures based on the details of the requirement described in the AI Guidelines for Business. Specific examples are described in 4.1.2 below. It should be noted that conformity or nonconformity with the AI Guidelines for Business does not immediately mean there was or was not negligence, since, as described in 2.1.3 above, the AI Guidelines for Business encourage business operators to make voluntary efforts and do not require compliance with everything contained in them, nor do they provide exhaustive descriptions of the measures required for each setting.

Even if an AI business user has taken appropriate measures in the context described in (A) and (B) above, AI may provide undesirable output. If any rights infringement or damage occurs as a result of a failure to correct the output, whether or not the AI business user has violated the duty to avoid the consequences can be an issue. However, in a setting where dependable/substitutive AI is operated, if the AI business user has appropriately built and operated a work process incorporating an AI system assuming that the system meets certain requirements, the user is considered to have fulfilled the reasonable duty of care that can be fulfilled as the user of the system, as with the ATM system described as an example in 2.2.2 above. In other words, in the case of dependable/substitutive AI, if the AI business user has taken appropriate measures in the context described in (A) and (B) above, the user will not be considered to have a duty to avoid consequences that extends to correcting all inappropriate AI output by examination or otherwise.

(3) Liability of AI developers and AI providers who sell or provide dependable/substitutive AI

As for AI developers and AI providers, the approach for determining their liability is itself no different from the one taken in the case of assistive/supportive AI, and the issue is whether they have taken the required design and explanatory measures. However, the required level would be higher than for assistive/supportive AI.

Specifically, in the case of dependable/substitutive AI, the issue is whether reasonably possible design measures and updates have been carried out to secure and maintain the standards and safety as described in (2) above, assuming that no final human judgment or action is involved in this case and the AI's output may directly cause rights infringement or damage (the duty of care in design). The necessary measures inevitably depend on the specific circumstances of each case; for example, key issues may include whether the AI's accuracy and safety have been

improved to a reasonably feasible extent during development, and whether reasonable safeguards have been considered assuming that undesirable outputs may occur within a certain range. Another important viewpoint is how dependable/substitutive AI should be used in the work process, as described in (2) above. AI developers and AI providers are expected to analyze, as necessary within the scope of what can be reasonably foreseen, and explain to AI business users important information on risk control, for example, the limitations of the AI's functions and performance, methods of use, important risks, as well as external sources of danger that are difficult to address by the AI system alone, situations in which risks may increase, and the scope within which human involvement is desirable (the duty of care in explanation).

As with the duty of care described in Section 2.2.1(3), the content of the duty of care in design and explanation required of AI developers and AI providers may vary depending on their respective roles within the value chain.

2.2.3 Summary on each category

Putting together the approaches described above, whether an AI falls under the category of assistive/supportive AI or that of dependable/substitutive AI should be determined by considering the following and other factors from the viewpoints described above: the functions of the AI and the setting where it is used; the accuracy and safety of the AI; the level of operational risk posed by AI; any specialized nature of the AI; and whether or not the AI is subject to relevant regulations. Determining the category of an AI would be difficult in some real cases. Chapters 3 and 4 of this Guidance discuss hypothetical cases to explain what AI falls under which category.

The figure below re-organizes the descriptions of, and the approaches to determine liability with respect to, assistive/supportive AI and dependable/substitutive AI.

		Assistive/supportive AI	Dependable/substitutive AI
Description		• This category of AI is used exclusively to **assist or support decision-making** and is intended to eventually involve human judgment or action. • As described in (A) through (C) below, there are cases where AI should be used exclusively as assistive/supportive AI. (A) Cases where the AI cannot be regarded as substituting human judgment in light of the AI's functions and the settings where it is used. (B) Cases where the final human judgment is required for reasons of regulatory laws. (C) Cases where the AI's output entails a potential risk of infringing third-party rights, and this requires human assessment or examination.	• This category of AI is provided on the assumption that it will substitute human judgment or action, and is expected to **be used by relying on the AI's output**. • The following two requirements must be met by this category of AI. (a) The AI is expected to bring benefits that are difficult to achieve if human judgment or action is involved (i.e., necessity). (b) The AI provides a certain level of accuracy and safety (i.e., accuracy and safety). →While the required level is different for different activities, if the AI's work level is equivalent to or higher than the level of work by a reasonable person in terms of accuracy and safety,* it would be found reasonable to leave judgment to the AI.
Approach to determine liability	Liability of AI business users	• An AI business user is required to **use the AI by exercising the due care that should properly be exercised by an AI business user** in the relevant situation.* • Specifically, the AI business user's duty of care may include, among other things: using the AI by assessing the correctness and appropriateness of the AI's output; collecting information necessary for such usage; and taking measures for the user's use that are necessary for such usage.	• Since the AI business user's judgment is not involved, the duty of care is transformed from the duty to make appropriate judgment or take appropriate action into the duty to **properly build and operate a work process incorporating the AI system**. • The AI business user will not be considered to have a duty to avoid consequences that extends to correcting all inappropriate AI output by examination or otherwise.
	Liability of AI developers/providers	• The assumption is that the AI developer/provider provides **explanations** on, among other things, the performance limits of, and important risks involved in, the AI in order to allow AI business users to examine the appropriateness of the AI's output and to correct the output. • The AI developer/provider may be required to take certain **design measures** with respect to risks that cannot be easily foreseen or dealt with by AI business users.	• The AI developer/provider is required to take reasonably possible **design measures** to secure and maintain safety as described above and to take **explanatory measures** such as analyzing important information on risk control and providing information to AI business users.

*The level of the duty of care required in each setting is determined based on, among other things, the size of the risk arising from the tort, the gravity of the infringed interests, and the attention of a reasonable person who is in the occupation or position. See note 32 above for approaches that may be taken in settings where it is difficult to directly compare the level of AI work and that of human work.

Chapter 3 Cases Involving Assistive/Supportive AI

This chapter discusses cases involving assistive/supportive AI. Currently, most AIs seem to fall under the category of assistive/supportive AI. As described in 2.2.1 (1) above, the substantial grounds for this include the following: (A) the AI cannot be regarded as substituting human judgment in light of the AI's functions and the settings where it is used; (B) the final human judgment is required for AIs for reasons of regulatory laws; and (C) AI's output entails a potential risk of infringing third-party rights, and this requires human assessment or examination. From these viewpoints, the following hypothetical cases 1 through 4 involve assistive/supportive AI.

In the case of assistive/supportive AI, the AI business user is required to use the AI's output to an appropriate extent by exercising the due care that should properly be exercised by an AI business user. Factors that are important in determining the user's duty of care may vary depending on the characteristics of the AI, the nature of the risks involved, and the attributes of the AI business user, among other things.

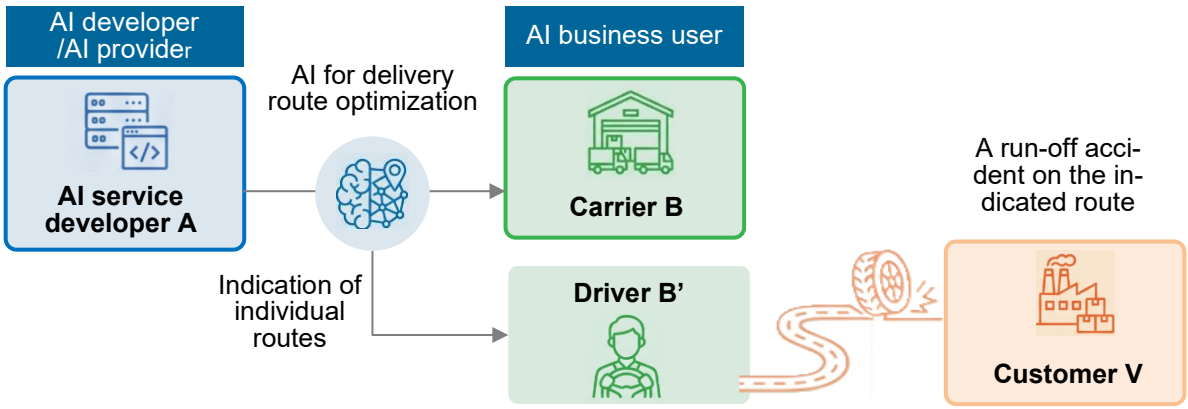
For instance, whether or not the AI's output is appropriate in the context of the AI business user's business purpose should naturally be assessed and examined by exercising the due care that should properly by exercised by the AI business user. The AI business user's duty of care includes using AI by considering whether or not the output is appropriate (hypothetical case 1 in 3.1 below and hypothetical case 2 in 3.2 below). From this viewpoint, it would be generally easy for the AI business user to control the appropriateness of the AI's output, and it is hard to imagine that the liability of the AI developer/provider will become an issue.

On the other hand, there are cases where the AI's output does not at a glance clearly indicate where risks lie (hypothetical case 3 in 3.3 and hypothetical case 4 in 3.4 below). Even in these cases, the AI business user should, as in other cases, use the AI's output to an appropriate extent in the context of the user's final judgment and action. However, the measures necessary for this, which are the focus of the user's duty of care, may be collecting certain information in advance or taking necessary advance measures for the user's use, rather than simply examining the AI's output itself. In addition, in these cases it is not always easy for the AI business user to specifically foresee the risks involved in the AI or to use the AI's output by controlling it, due to which it becomes an issue whether the AI developer/provider should have taken certain design measures.

3.1 Hypothetical case 1: AI for delivery route optimization

Carrier B determines daily delivery routes by using the AI system for delivery route optimization provided by company A,[35] which is engaged in the business of developing and providing AI. When delivery destinations, vehicle information, time limits, and other conditions are entered in the system, the system can automatically create a vehicle allocation and routing plan that has been optimized in terms of efficiency and safety by taking into account such conditions as positional relationships and distances between delivery destinations, appointed times for delivery, and traffic conditions. This allows cost reduction, such as reduction in time required for delivery and reduction in the number of vehicles that need to be assigned. Company A had explained to carrier B these functions and characteristics of the AI and the requirement that drivers must check the safety of the final road condition, such as whether the road width and surface and visibility that allow for safe travel have been ensured. In order to make transportation more efficient, carrier B had instructed its drivers to follow, as a rule, the routes output by the AI. On the other hand, carrier B trained and instructed its employees to check driving safety on-site and, if safety cannot be secured, to drive along safe routes on their own judgment.

One day, when carrier B's driver B' was performing transportation service, a bad, narrow road unfit for large vehicle passage was shown by the AI as the "optimal route." Driver B' hesitated to enter the road, but followed the AI's output by taking into consideration, among other factors, coordination with the overall delivery plan and delay penalties, since the vehicle allocation schedule for the day was extremely tight. As a result, the vehicle ran off the road due to a lack of sufficient space for passage. This resulted in a significant delay in delivery. At the same time, the resulting impact caused damage to the delivery to customer V, who suffered damage in order to procure replacements and take other action.



[35] Since company A provides carrier B with services using the AI system developed by company A, company A is in the position of both an AI developer and an AI provider (see the AI Guidelines for Business, p. 5 and the AI Guidelines for Business Appendix, pp. 9 - 10).

In the above case, the issue was whether or not driver B' was liable (under Article 709 of the Civil Code) for the accident that resulted from following the route indicated by the AI. If driver B' is held liable, carrier B would be held liable as the employer of driver B' (Article 715 of the Civil Code).

As for company A, the issue was whether or not it is held liable in any way for the fact that the route indicated by the AI turned out to be unfit for vehicle passage.

3.1.1 The position of AI in hypothetical case 1

Map apps and other route guidance services that indicate the shortest route to the destination have long been provided. With the progress of AI technology, it has become possible to present optimized routes by analyzing a wider variety of factors. Some of these AIs provide route guidance by taking the safety of driving routes also into account by taking into safety information into consideration, such as regulations, the existence of narrow roads, map information, and speed information, as well as information on delivery efficiency such as positional relationships and distances between delivery destinations and appointed times of delivery.[36]。

Thus, a route optimization AI falls under the category of assistive/supportive AI because, while a route optimization AI has functions to improve safety in some way, it does not have functions to provide safety judgment in individual settings. In settings like hypothetical case 1, the factor that is most directly linked to the avoidance of accidents is the road conditions, such as the road width and surface, and the driver's recognition of and judgment on them, as well as other information obtained on the actual road and information on the actual operation. In order to prevent accidents, the driver is required to make safety judgment based on these types of information. The above AI calculates routes not by taking these types of information into account but based on vehicle information, road information, and other external information. Even if the AI is designed to ensure safety to some extent, it cannot be regarded as perceiving the situation or making judgments for the driver who should do these activities. Therefore, the AI falls under the category of assistive/supportive AI whose output should only be used as reference information.

3.1.2 Liability of carrier B and driver B' as AI business users

As described above, a route optimization AI should be used (as assistive/supportive AI) to assist the AI business user's judgment. The fact that driver B' adopted the AI's output does not constitute a factor that changes the level of driver B's duty of care to drive safely.

Driver B's negligence may be denied in cases where there were exceptional circumstances that made it reasonable even for a reasonable driver engaging in transportation service to drive along

[36] See: Policy Research Institute for Land, Infrastructure, Transport and Tourism (2023). "Chiiki no Yuso Shigen wo Katsuyo shita Butsuryu Network Saiteki-ka ni kansuru Chosa Kenkyu (Saishu Houkoku) [Research Study on Logistics Network Optimization Utilizing Local Transportation Resources (Final Report)]" (in Japanese). *PRILIT Report No. 172*. pp. 12 - 14.

the road where the accident occurred because there was a situation where the danger of the road was very hard to be detected. On the other hand, it is generally easy to become aware of the danger of driving a large vehicle along a narrow road, and the driver can normally take action such as avoiding accidents by turning back if necessary. Therefore, the driver is more likely to be found to be at fault unless exceptional circumstances exist.

If driver B' is found to be at fault (Article 709 of the Civil Code), carrier B is held liable as the employer of driver B' (Article 715 of the Civil Code).

There is also a possibility that if carrier B had instructed its drivers to always follow the output of the above AI despite the fact that the AI falls under the category of assistive/supportive AI in light of the AI's functions as described in 3.1.1 above, this contributed to the occurrence of the accident. If, unlike in hypothetical case 1, carrier B had instructed its drivers to always follow the AI's judgment in order to keep the company-wide delivery schedule or for such other reasons, there is also a possibility that the employees of carrier B, who gave such instructions, may be found to be at fault.

3.1.3 Liability of company A as the AI developer/provider

The liability of company A as the AI developer/provider may become an issue in two different settings described below. However, in light of the facts involved in hypothetical case 1, it is fundamentally difficult to anticipate that company A would be held liable in either setting.

(1) A setting where carrier B compensated customer V for the damage and claims reimbursement or the like from company A

Assuming that company A and carrier B had entered into a contract for the provision of the route optimization AI, carrier B may claim reimbursement or damages from company A based on the contract and on other potential grounds, after compensating customer V for the damage incurred by it. The Study Group refrains from discussing this issue, since whether or not such a claim will be granted is up to the contract between company A and carrier B, (see 2.1 above). Important factors include, among others: what agreements were made on the performance and quality of the AI; what liability company A will have under the contract if the AI fails to provide the performance, quality, etc. intended under the contract; and what disclaimers were contained in the contract.[37]

[37] The opinion was expressed at a meeting of the Study Group that, in settings like hypothetical case 1, the contract in most cases contain disclaimers that usually limit company A's liability for damages and the like to cases involving intent or gross negligence on the part of company A. Among these cases, there would be very limited cases where company A's intent or gross negligence is found to have been involved in, and company A is held liable for, the occurrence of the accident in the setting of hypothetical case 1.

(2) A setting where customer V seeks damages directly from company A

In addition to (1) above, there can be an unlikely setting where customer V claims damages directly from company A (in addition to claims against carrier B). In this setting, customer V's claim against company A is based on tort liability as there is no contractual relationship between company A and customer V.

As described in 2.2.1 above, any undesirable output of an AI falling under the category of assistive/supportive AI is expected to be examined and corrected eventually by the AI business user's judgment or action. Therefore, company A is held liable for a tort in limited cases only. Company A had explained to carrier B the functions and characteristics of the AI and the requirement that drivers must check the safety of the final road condition, such as whether the road width and surface and visibility that allow for safe travel have been ensured. Therefore, it is unlikely that company A is found to have failed to provide necessary explanations. In addition, as long as these explanations have been provided, it is expected that, in the context of division of roles between the parties, the driver who actually drives along roads will choose safe driving routes on their own judgment. Therefore, it is generally difficult to imagine that company A is further required to take design measures.[38]

3.2 Hypothetical case 2: AI for assisting legal services

LLM provider C[39] provides a general-purpose large language model developed by it. AI service developer D builds and provides chatbot AI services that assist legal services by: fine-tuning LLM provider C's model to improve the model's legal reasoning ability; and searching relevant information from databases of laws and regulations and judicial precedents; and integrating retrieved information into model input (Retrieval-Augmented Generation; "RAG"). This AI service has the functions of displaying highly relevant literature and judicial precedents and generating reports that summarize legal analyses, in response to prompts that are entered by an attorney, who is an AI business user, and that consist of subjects being searched or specific cases. Attorney E has introduced the AI service into their services.

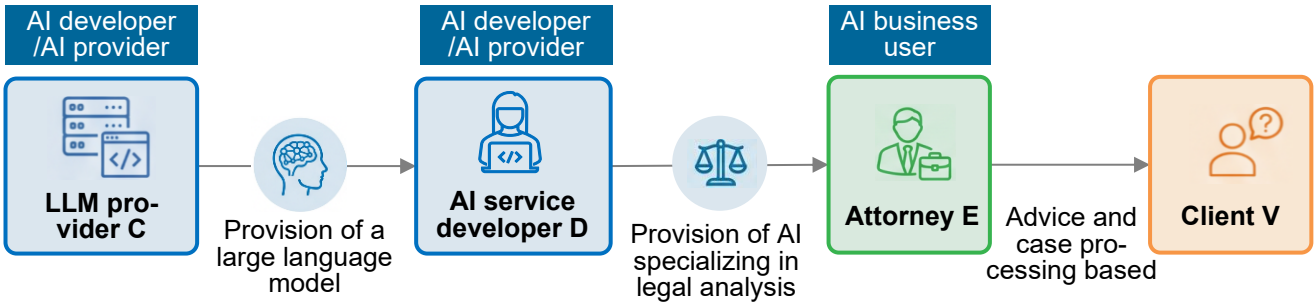
Attorney E entered, in the AI service, documents containing the history of a certain dispute on which Attorney E had been consulted by client V, as well as documentary evidence for the dispute, and requested an analysis addressing legal issues involved in the case and prospects for the case. According to the AI's output, there were a number of judicial precedents supporting the position of the client V, and the conclusion was that it was likely that the client's claims

[38] In a classroom case where company A provided services knowing that the AI has a strong tendency to guide drivers into dangerous roads, company A would be held liable for violating a duty to take design measures.

[39] Since LLM provider C provides a large language model developed by it to AI service developer D, LLM provider C is in the position of both an AI developer and an AI provider of the large language model. As for AI service developer D, since it provides the AI for assisting legal services developed by it to attorney E, AI service developer D is in the position of both an AI developer and an AI provider of the AI service.

would be granted. As this analysis agreed with attorney E's own experience and view, attorney E trusted the AI's output without personally researching judicial precedents and established a policy to maximize profits without considering a settlement. However, the AI actually cited a number of fictitious judicial precedents that did not exist. Most of the judicial precedents that existed did not support client V's claims.

Eventually, client V's claims were totally dismissed by the court, and client V suffered damage, including attorneys' fees for conducting the litigation.



In the above case, the issue is whether or not attorney E is liable for violating the duty of care of a prudent manager to client V under the retainer agreement (Article 415 of the Civil Code) and/or has general tort liability (Article 709 of the Civil Code), for processing the case by relying on the objectively wrong output of the AI without conducting their own necessary research on judicial precedents that support the client's position.

As for LLM provider C and AI service developer D, the issue is whether they are held liable in any way for the fact that the AI's output turned out to be wrong.

3.2.1 The position of AI in hypothetical case 2

In hypothetical case 2, the AI for legal analysis has the function of generating analysis regarding the application of laws and regulations based on the documents and evidence for individual cases, while consulting laws and regulations, judicial precedents, etc. The AI's output has an aspect that is similar in function and nature to the recognition and judgment of an attorney in the performance of their services.

However, an industry regulated by law requires considerations that are different from those in hypothetical case 1. While there are no specific regulations on attorneys' use of AI in their services, Article 72 of the Attorneys Act provides as follows: "No person other than an attorney or a legal professional corporation may, for the purpose of earning compensation, engage in the legal services such as provision of an expert opinion, representation, mediation, or settlement of a litigation case ... or other general legal cases, or may engage in mediation services related to these cases." If an AI developer/provider who is not an attorney or a legal professional corporation provides an

AI without the involvement of an attorney in a manner set forth in Article 72 of the Attorneys Act, this may constitute a violation of the same article.

Based on Article 72 of the Attorneys Act described above, among other things, hypothetical case 2 assumes that attorney E is provided with assistive/supportive AI that is used only to assist or support the attorney in making judgments.

3.2.2 Liability of attorney E as an AI business user

While Attorney E owes the duty of care of a prudent manager under the retainer agreement and the duty of care under tort law, there would be substantially no difference in the framework for determining negligence between contract law and tort law.[40] Since the AI in question in hypothetical case 2 falls under assistive/supportive AI as described in 3.2.1 above, whether or not attorney E is found to have violated the duty of care of a prudent manager or have been negligent is determined depending on whether or not the attorney's action based on the actual process of case processing was reasonable. The fact that attorney E used an AI neither raises nor lowers the level of the attorney's duty of care.

What must be kept in mind when determining the level of duty of care owed by attorney E is that the attorney owes the duty of care as a professional attorney. When determining negligence under tort law, the standard for determination is not the care exercised by a notional ordinary person but the care exercised by a reasonable person described for each role in social life, such as the defendant's occupation, status, or position. Attorney E owes the duty of care as an attorney who is able to exercise such care.[41]

Therefore, in the context of whether or not attorney E was negligent, the question is what action would have been taken by an attorney who is able to exercise ordinary care and whether or not attorney E's action was in line with that action. If attorney E used an AI, the issue is whether or not attorney E's judgment made by consulting the AI's output was in line with the duty of care as an attorney. In the context of hypothetical case 2, whether or not judicial precedents supporting the client's position really exist is very important information in processing the case. Any attorney would have to meticulously research and confirm the existence of such precedents, and it is unlikely that the liability of attorney E, who relied only upon the AI's output and failed to conduct sufficient research and confirmation, would be denied.

---

[40] A mandatary bears a duty to administer the mandated business with the due care of a prudent manager in compliance with the main purport of the mandate (Article 644 of the Civil Code). The "due care of a prudent manager" is specifically determined by the level of education or expert knowledge required to administer the intended business and by the abilities and qualities of the mandatary that the mandator knows or should know. In particular, if the mandatary is a professional business administrator, the mandatary is subject to such level of due care of a prudent manager as is generally expected of persons in the same occupation or position (Yamamoto, Yutaka (ed.) (2018). *Shin Chushaku Minpo [Japanese civil law annotated, new edition] (14) Saiken [Claims] (7) (in Japanese).* p. 251 [Ichiki, Takayuki]. Yuhikaku). This is similar to the framework for determining negligence described in 2.1.1 above in that the general level of care objectively required in the relevant business becomes an issue.

[41] Kubota, Atsumi (ed.) (2024). *Shin Chushaku Minpo [Japanese civil law annotated, new edition] (15) Saiken [Claims] (8) (in Japanese).* pp. 366-368 [Hashimoto, Yoshiyuki]. Yuhikaku.

In this regard, the following opinion was expressed at a meeting of the Study Group in light of the fact that settings have been emerging where AI's performance is partially equal to or higher than professionals in making judgments: based on the cognitive scientific finding that human attention is suggested to decline when a person works in coordination with advanced AI, there may be settings where it is not desirable to immediately find a professional to be negligent on the grounds that they were unable to detect an AI's wrong output, and it is important to discuss the framework for determining liability when this is the case (for instance, an approach where a professional's duty of care includes judging how AI should be utilized in the relevant professional area, or making sophisticated judgments on the appropriate usage of AI).

3.2.3 Liability of LLM provider C and AI service developer D as the AI developer/provider

The liability of LLM provider C and AI service developer D, which developed and/or provided the AI, would be the same as in hypothetical case 1 (3.1.3 above). Specifically, the possible settings are: (1) attorney E has compensated client V for the damage and claims reimbursement or damages from LLM provider C and/or AI service developer D[42]; and (2) client V claims damages directly from LLM provider C and/or AI service developer D (in addition to claims against attorney E).[43] In either case, there would only be a limited possibility that LLM provider C and/or AI service developer D is held liable.

In setting (2) above, where the claims are based on tort law, potential issues are: (A) whether LLM provider C and/or AI service developer D provided the AI business user with explanations on the performance limits of the service provided by it and on how to use the service; and (B) whether LLM provider C and/or AI service developer D is considered to have provoked wrong handling of the dispute by the attorney who was an AI business user. In the context of (A), it would be important that LLM provider C provided AI service provider D with clear explanations on the performance limits of the general-purpose large language model provided by LLM provider C to AI service provider D, how to use the model, and important risks involved in the model, among other things, and that AI service provider D provided attorney E with corresponding clear explanations on the RAG service provided by AI service provider D to attorney E. As long as these explanations

[42] As with Section 3.1.3(1) above, although it depends on the specific contractual terms, important factors include, among others: what agreements were made on the performance and quality of the AI; what liability company A will have under the contract if the AI fails to provide the performance, quality, etc. intended under the contract; and what disclaimers were contained in the contract.

[43] The opinion was expressed at a meeting of the Study Group that, in hypothetical case 2, it is not easy for client V to identify the internal circumstances involved in the development and provision of the AI service, such as the contractual relationships between LLM provider C, AI service developer D, and attorney E, or whether or not sufficient explanations were provided, due to which client V would usually blame attorney E, who has a direct contractual relationship with client V, in which case the liability of LLM provider C and AI service developer D will only be an issue in terms of reimbursement claimed by attorney E from AI service developer D or by AI service developer D from LLM provider C, mostly after attorney E has provided compensation to client V.

have been provided, the attorney is expected, in the context of division of roles between the parties, to eventually make a judgment by taking the AI's output into account on the attorney's own responsibility. Therefore, it is generally difficult to imagine that LLM provider C or AI service provider D is held liable.

3.3 Hypothetical case 3: Image generation AI

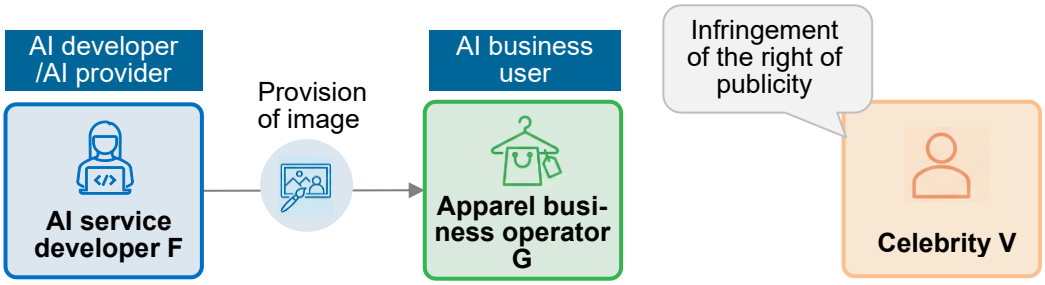
[Case (a)] In making its ads and other materials, apparel business operator G used an image generation AI developed and provided by AI service developer F. From the images generated by the AI, apparel business operator G's person in charge G' chose some images that match the atmosphere of the company's brand and used them in web ads for the company's goods. The image used in the main ad closely resembled V, who was well known nationwide, but person in charge G' was not aware of this. Person in charge G' had no intention to generate images that resemble celebrities, including talent V. When entering prompts, person in charge G' did not enter any proper nouns but generated patterns of a wide range of persons and poses. V found that images closely resembling them were used in ads, and claimed that their right of publicity had been infringed.

The image generation AI developed and provided by AI service developer F is a general-purpose service intended to meet a wide variety of needs, and provided images such as realistic landscapes or person images. The terms of use of the AI allowed generated products to be used for commercial purposes. The AI's products were used in creative activities, icons used on SNS, and corporate advertising, among other things. The data set used by AI service provider F in developing the AI was a public data set containing a large number of URLs that display approximately 1 billion images, including images and the like of persons. Since the data set was created by web crawling, it was difficult to check if images of talent V were contained in the data set. However, the data set contained no data labeled with celebrities' names or any similar data. In addition, re-captioning, training in latent space, or such other measures had been taken to prevent original data from being reflected in generated images without change. During the model assessment conducted by AI service developer F or in any post-release feedback received from users, there was absolutely no repot that images closely resembling talent V were generated.

[Case (b)] In making ads and other materials, apparel business operator G's person in charge G' used an image generation AI developed and provided by AI service developer F. G' became aware that images generated by the AI included images that closely resembled talent V. Person in charge G' used these images in web ads for the company's goods with the intention of promoting sales of the company's goods by taking advantage of talent V's name recognition. Talent V found this and claimed that their right of publicity had been infringed.

All other facts, such as the model training methods used by the image generation AI and the provisions of the terms of use of the AI, are the same as described in the second paragraph of case (a).

[Case (c)] AI service developer F developed an AI model that, after being trained using a data set consisting of image data of celebrities labeled with their names, frequently generates images identical or similar to these celebrities, including talent V, in response to input prompts consisting of the celebrities' names. AI service developer F sold to apparel business operator G and other customers image generation AIs that used the AI model, featuring it as an AI intended to generate various images of celebrities.



In the above case, the issue is whether or not person in charge G' is held liable under general tort liability (Article 709 of the Civil Code) for using images that closely resembled talent V in apparel business operator G's ads. If person in charge G' is held liable, apparel business operator G is held liable as the employer of person in charge G' (Article 715 of the Civil Code).

As for AI service developer F, the issue is whether or not it is held liable for aiding and abetting person in charge G's activities (Article 719, paragraph (2) of the Civil Code) if the person in charge intentionally infringed talent V's right of publicity. Furthermore, in exceptional cases where, due to the manner in which AI service developer F acted, the AI service developer is considered to have played a leading role in infringing the right of publicity, the liability of AI service provider F may also become an issue (Article 709 of the Civil Code).

3.3.1 The right of publicity and image generation AI

The right of publicity is the right of a person whose images or the like[44] have customer attraction to exclusively utilize the economic value and interests arising from the customer attraction. The existence and details of right of publicity were recognized by reference judicial precedent 4 below. This judicial precedent listed three types of activities ((A) through (C) below) (hereinafter referred to as the "Three Types") that are found illegal under tort law due to infringement of the right of publicity.

[44] "Images or the like" of a person means the person's personal identification information including, for instance, the person's signature, voice, pseudonym, and stage name (Nakashima, Motoyuki (2015). *Hankai [Case commentaries], Saihankai Minji Hen Heisei 24-nendo (jou) [Commentaries on supreme court judgments on civil cases in fiscal 2012 (vol. 1)]*. p. 41).

Reference judicial precedent 4: Supreme Court Judgment, February 2, 2012, *Minshu (Supreme Court Civil Case Reports), Vol. 66, No. 2,* p. 89.

Nature of the right

- **The name, images, etc. of a person symbolize the person, and the person has the right, derived from the moral right, to be protected from abuse of these pieces of information.**
- **There are cases where a person's images or the like have customer attraction that promotes sales of goods or other benefits. The right to exclusively use this customer attraction (the "right of publicity") is based on the commercial value of the images or the like themselves. Therefore, the right of publicity can be regarded as constituting one of the rights derived from the aforementioned moral right.**
- **A person whose images or the like have customer attraction may face situations where their images or the like are used in news reports, editorials, creative works, etc. due to social attention drawn to them. In some cases, the person should endure such use as legitimate acts of expression.**

Criteria for determining rights infringement

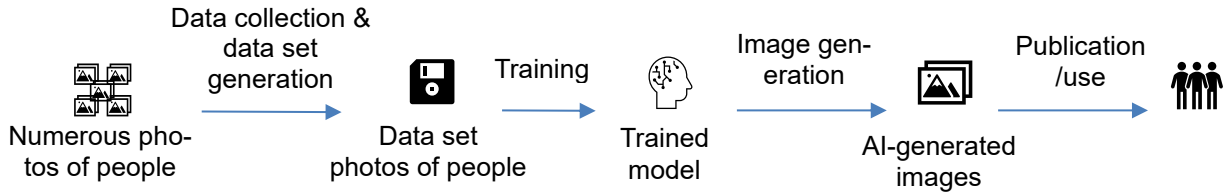
- **Using a person's images or the like without permission is illegal under tort law as an infringement of the person's right of publicity in cases where the dominant purpose of such use is regarded as utilizing the customer attraction of the images or the like, such as where: (A) the person's images or the like are used as goods or the like[45] that are independent objects of appreciation; (B) the person's images or the like are attached to goods or the like with the intention of differentiating the goods or the like; or (C) the person's images or the like are used in ads for goods or the like.**

The typical setting where the right of publicity becomes an issue used to be where a specific person is photographed and the person's images are directly used in ads or the like. This was also the case with reference judicial precedent 4 described above, where the dispute was whether publishing photographs of a singer in a magazine without the consent of the singer infringed the singer's right of publicity.

On the other hand, in the case of image generation by generation AI, the process proceeds as follows: first, a data set for training is generated by collecting data on images of a large number of persons and the like; an AI model is trained by using the data set; images of persons are generated by entering instructions (prompts) to generate images into the trained model; and the images are used in ads or the like. In the case of image generation by AI, numerous images are contained in the data set for training. In the process of training an AI model, those data have been converted into parameters, which means that photographs of people are not directly used in the generation of images.

[45] "Goods or the like" means items commercialized as goods or services. If the use of images or the like is for private purposes only, the use is considered not illegal. Therefore, it is considered that images or the like constitute goods or the like only if they are used in the course of business (i.e, commercially used) (note 44 above, Nakashima, p. 40).

This structure generates the issue of how to determine whether the right has been infringed and whether there was negligence in cases where an image generation AI accidentally generated images that resemble a specific celebrity and where these images were eventually used.[46]



3.3.2 The position of AI in hypothetical case 3

An image generation AI like the one involved in hypothetical case 3 has the function of generating images according to prompts. These images are intended to be used at the discretion of the AI business user after additional processing and the like if necessary. This may result in infringement of a third-party's right of publicity depending on the nature of the images generated. This would require human assessment and examination in the process of utilization in order to prevent infringing images from being used externally. In this sense also, this type of AI is not fit for use as dependable/substitutive AI. In light of the above discussions, the AI described above is another type of AI that falls under assistive/supportive AI.

If the output of an assistive/supportive AI includes something that infringes third-party rights, the AI business user is required to use the AI's output to an appropriate extent at the user's own judgment.[47] However, it is generally difficult for the AI business user to check each of the images output by AI against images of numerous right holders. It is possible that the AI business user is unable to give sufficient consideration to right holders, resulting in infringement of their rights of publicity. In such case, the AI developer/provider may be required to take measures to prevent rights infringement subject to certain requirements (3.3.4 below).

---

[46] While the following discusses issues related to liability of each party for tort damages, another potential issue is whether or not injunctive relief can be granted. The requirements for granting injunctive relief generally include: (A) infringement of rights or legal interests; (B) illegality; and (C) a reasonable probability of the occurrence of substantial infringement (i.e., a causal relationship) (Otsuka, Tadashi (ed.) (2022). *Shin Chushaku Minpo [Japanese civil law annotated, new edition] (16) Saiken [Claims] (9) (in Japanese)*. p. 429 [Otsuka, Tadashi]. Yuhikaku). With regard to the requirement of illegality listed in (ii) above, Nakashima (note 44 above, pp. 69 - 70) mentions the view that being found illegal under tort law is not sufficient to be granted injunctive relief, in light of the general view that when the right to demand injunctive relief is exercised on the grounds of infringement of moral rights, the level of illegality required for granting injunctive relief is higher than that for granting damages. However, Nakashima presents his view that, in conclusion, it is reasonable to forthwith grant injunctive relief if a tort is established. On the other hand, the view has been expressed that criticizes the view that injunctive relief should be granted for threatened infringement of the right of publicity, even without express provisions, in accordance with the same criteria as those that apply to tort damages (Saito, Kunifumi (2023). *Privacy to Shimei/Shozo no Hou-teki Hogo [Legal protection of privacy and names/images]*, p. 236. Nippon Hyoron Sha).

[47] With respect to measures to be taken by the AI business user in relation to the right of publicity, the AI Guidelines for Business make no direct mention but expect the AI business users to make appropriate use by giving consideration to safety to prevent harm to the property and other benefits of stakeholders. See "<Reference> Measures related to the right of publicity expected under the AI Guidelines for Business" on p. 35 below.

3.3.3 Case (a)

(1) Liability of apparel business operator G and its person in charge G' as AI business users

In case (a), apparel business operator G and its person in charge G' accidentally used images closely resembling talent V in the business operator's web ads.

As described above, apparel business operator G and its person in charge G' are required to use generated images to an appropriate extent, and whether or not they are liable is determined according to the conventional requirements for claiming damages for infringement of the right of publicity. Judicial precedents issued after reference judicial precedent 4 and relating to these requirements include those that discuss rights infringement and intent/negligence as one issue[48] and those that discuss these subjects separately.[49] In settings where an image generation AI is used, the issue of rights infringement, i.e., whether or not the use and other handling of the generated images constitute an infringement of the right of publicity based on the legal assessment described in 3.3.1 above, and the issue of intent/negligence, i.e., whether or not the AI business user owed the duty to prevent the rights infringement, may be different from each other in terms of quality. Therefore, these issues are organized separately below.

(A) Rights infringement

As described above, infringement of the right of publicity is considered to occur in cases where images or the like are used and where the dominant purpose of such use is utilizing the customer attraction of such images or the like, such as in cases where the user's usage falls under any of the Three Types.

First, in cases where images generated by an image generation AI are used, it may become an issue whether or not these images constitute "images or the like" in the first place, since photographs or the like of the relevant person are not directly used. In reference judicial precedent 5 described below, the court denied infringement of the right of publicity by the use of drawings of a celebrity, on the grounds, among others, that the drawings did not resemble the celebrity enough for people who know him to easily identify that the drawings were of him. In light of this judicial precedent, even if the relevant person's images themselves are not used, generated images would potentially constitute "images or the like" if the generated images resemble the person

[48] For instance, Intellectual Property High Court Judgment, October 16, 2013 (unreported); Osaka High Court Judgment, November 16, 2017, *Hanrei Jiho, No. 2409,* p. 99; and Intellectual Property High Court Judgment, June 30, 2025 (unreported).

[49] For instance, Tokyo District Court Judgment, April 26, 2013, *Hanrei Times, No. 1416,* p. 276; Intellectual Property High Court Judgment, February 20, 2020 (unreported); and Intellectual Property High Court Judgment, September 13, 2023 (unreported).

enough for people who know that person to easily identify that the generated images are of the person.[50]

Reference judicial precedent 5: Tokyo District Court Judgment, June 14, 2005, Hanrei Times, No. 1217, p. 310.

- **In a case where graphics of a person who resembled the plaintiff, a rock singer, were used on pachinko machines and where the dispute was over an alleged infringement of the right of publicity of the plaintiff, the court denied the alleged infringement of the plaintiff's right of publicity on the grounds, among others, that while it is inferred that the graphics of a person were inspired by the image of the plaintiff, and even though the graphics objectively resemble the plaintiff to some extent, it is difficult to find that the graphics resemble the plaintiff enough for people who know the plaintiff to easily identify that the graphics are of the plaintiff.**

Then, when an image generation AI is used, there may be cases where images or the like that resemble the relevant person are accidently generated and where the AI business user does not necessarily intend to infringe the person's right of publicity. The point of consideration would be whether or not the AI business user has the dominant purpose of utilizing the customer attraction of the images or the like, when various factors are taken into account. Judicial precedents issued after reference judicial precedent 4 take into account objective circumstances associated mainly with the utilization of the customer attraction of the images or the like when considering the alleged rights infringement, such as how the images or the like were used or published,[51] the volume and content of the text attached to the images or the like,[52] and the purposes and natures of the media in which the images or the like were published.[53]

In case (a), the following factors are considered to strongly support the inference that the dominant purpose was to utilize the customer attraction of talent V's images or the like, and would fall under type (iii) of the Three Types, i.e., using images or the like in ads for goods or the like: the

[50] Nakashima (note 44 above, p. 41) argues that, in case of character goods with designs of an animal or the like that resemble a celebrity, if the designs are well known to consumers as an identifier of the celebrity, the designs constitute images or the like since they are deemed to constitute information for personal identification of the celebrity. According to the explanation of the above argument, even in case of designs of an animal or the like that are more abstracted than graphics of a person, those designs may constitute images or the like depending on how well known the designs are as an identifier of the relevant person.

[51] Reference judicial precedent 4 above. It has been pointed out that, while the judicial precedent lists the Three Types as examples in the context of criteria for an infringement of the right of publicity, the court does not discuss whether or not the relevant use falls under any of the Three Types in their application to the case but discusses whether or not the dominant purpose of such use is regarded as utilizing the customer attraction of the relevant celebrity, by analyzing the purpose, method, and manner of the use (Ishii, Mio (2018), "Publicity ken to Hyogen no Jiyu [The right of publicity and freedom of expression]" (in Japanese). *Nihon Daigaku Chizai Journal [Journal of Intellectual Property, Nihon University], Vol. 9,* p. 58).

[52] The Intellectual Property High Court Judgment dated October 16, 2013, referred to in note 48 above (unreported), takes into account, among other factors, how the photos were published in the book and whether the document attached to the photos has an independent significance.

[53] In the note 49 above, Intellectual Property High Court Judgment of February 20, 2020 (unreported), the court noted that the purpose of the website on which the images or the like were published was the advertising and promotion of sales of goods or the like. In the note 49 above, Intellectual Property High Court Judgment of September 13, 2023 (unreported), the court noted that the website directed users to goods sales sites and fan club sites, and played a role in encouraging purchases of goods and membership in fan club sites.

fact that talent V is a nationwide celebrity; the fact that the generated images closely resemble talent V; and the fact that apparel business operator G used the generated images in its main ad to promote sales of its goods. Taking these factors into account, even if person in charge G' subjectively had no intention to infringe talent V's right of publicity, there is a high possibility that the use would be assessed as objectively having the dominant purpose of utilizing the customer attraction of talent V's images.[54]

On the other hand, unlike in case (a), possible cases where the purpose of utilizing customer attraction may be denied include the following. In such cases, if it is assessed that the dominant purpose was not to utilize customer attraction, rights infringement would be denied.

- Cases where the generated images resemble a local talent who is only recognized in certain regions, and it is difficult to assess that there was an intent to utilize the customer attraction of that talent in light of the purpose of the ads and other factors.
- Cases where the medium in which the generated images are published is not intended for promotion of sales of goods.
- Cases where the importance of the generated images in the medium in which they are published is low.

(B) Intent/Negligence

Judicial precedents that determine intent/negligence separately from rights infringement take into account, in determining intent/negligence, the prior circumstances relating to the use of images[55] or the like and the nature of the actor[56].

Given that the image generation AI involved in case (a) falls under assistive/supportive AI, the AI business user is considered to have the duty to investigate and confirm, within a reasonable scope, whether using such images in ads infringes the right of publicity. As an entity that creates ads to be published to the general public, apparel business operator G has a duty of care to investigate and confirm with reasonable due diligence. In particular, given that the generated

---

[54] If the AI business user is not aware that the generated images resemble a real person whose images or the like have customer attraction, there is a possible view that this lack of awareness itself negates the "dominant purpose of utilizing customer attraction." However, in light of judicial precedents that focus on objective factors such as the manner in which images or the like are used or published when determining rights infringement, it is not immediately the case that rights infringement is denied merely because the user was not aware of the resemblance to a real person. Rather, it is considered that whether or not there was an "dominant purpose of utilizing customer attraction" would be determined by taking into account circumstances such as the degree of resemblance and the manner of use.

[55] In the note 49 above, Tokyo District Court Judgment of April 26, 2013, *Hanrei Times, No. 1416,* p. 276, the court, in determining intent/negligence, noted, among other things, that the defendant had received complaints regarding infringement of the right of publicity even before the publication of the magazine.

[56] In the note 49 above, Intellectual Property High Court Judgment of September 13, 2023 (unreported), the court, in determining intent/negligence, noted that the defendant, as a management company for an artist, was in the position of being required to carry out its business with appropriate care with respect to the handling of the right of publicity held by the artist, and rejected the defendant's arguments regarding intent/negligence.

images closely resembled a nationwide celebrity, the possibility is high that negligence would be found on the grounds that it was possible to avoid using the images through simple investigation and confirmation.

On the other hand, unlike in case (a), in cases such as where AI generated images resembling a local talent who is only recognized in certain regions during the course of use of the AI, where the resemblance was not intended by apparel business operator G or person in charge G' but was only accidentally generated[57], and where it would have been difficult to specifically foresee infringement of the right of publicity even if apparel business operator G had exercised reasonable due diligence, there is a possibility that negligence may be denied.

(2) Liability of AI service developer F as the AI developer/provider

The issue in the context of AI service developer F's provision of the image generation AI[58] is whether or not F is held liable for aiding and abetting the tort committed by apparel business operator G and its person in charge G', and whether or not AI service developer F itself is found to have played a leading role in the infringement of the right of publicity.

However, case (a) involves the issue of whether or not the AI business user negligently infringed the right of publicity. The liability for aiding and abetting another person's tort is normally an issue in cases like case (b) described in 3.3.4 below, where the infringer committed an intentional tort. In conventional judicial precedents as well, liability for aiding and abetting has basically been an issue in relation to intentional torts (see the note 17 above).

Furthermore, AI service developer F itself is considered to have played a leading role in the infringement of the right of publicity only in exceptional cases such as case (c) described in 3.3.5 below.

For these reasons, F's liability is unlikely to be found in case (a), and will only be discussed in cases (b) and (c) below.

---

[57] Circumstances that may demonstrate that apparel business operator G and/or person in charge G' did not intend to generate images resembling a celebrity and that such images were generated only accidentally include, for example: the fact that the AI in question has been equipped with rights infringement preventive measures such as those described in [3.3.4] below, and that G and/or G' had confirmed that the possibility of real celebrities being generated was low; and the fact that G and/or G' managed the use of prompts so that proper nouns for real celebrities were not entered at the time of use, among other things.

[58] If images of talent V were included in F's training data, whether this itself constitutes an infringement of the right of publicity may also become an issue. However, as described in 3.3.1 above, images in the data set are not directly used in output or used externally, but are used only for adjusting the parameters of the AI model. Therefore, even if images of talent V were included in the training data, it is considered that this type of use would not constitute a use "for the dominant purpose of utilizing the customer attraction of the images or the like," and would not infringe the right of publicity.

3.3.4 Case (b)

(1) Liability of apparel business operator G and its person in charge G' as AI business users

Case (b) involves a setting where AI business user G' intentionally infringed talent V's right of publicity. In light of the approach described in 3.3.3 above, an intentional tort is established and liability for damages is found.

(2) Liability of F as the AI developer/provider

In the above case (1), AI developer/provider F's image generation AI was used in person in charge G's activities. The issue is whether or not AI service developer F is held liable for aiding and abetting (Article 719, paragraph (2) of the Civil Code) on the grounds that the provision of the AI facilitated person in charge G's tort.

Reference judicial precedent 6 below is particularly relevant to F's liability for aiding and abetting. In that judicial precedent, on the grounds, among others, that leasing karaoke equipment created a high probability of infringement of the copyrights by operators of establishments that used the equipment, the court imposed on the karaoke equipment lessor a duty under the rules of reason to deliver the karaoke equipment after confirming the conclusion of or application for a licensing agreement for the use of copyrighted works, and found the lessor jointly liable for the tort on the grounds that it had failed to fulfill the duty of care. As the grounds for this ruling, the judicial precedent cited the following as grounds for affirming a duty of care under the general principles of law: (i) the probability that leasing karaoke equipment would cause rights infringement; (ii) the gravity of the damage caused to the legal interests; (iii) the social position of the lessor; (iv) foreseeability; and (v) the possibility and ease of avoiding the consequences. [59]

Thus, when determining whether or not there is liability for aiding and abetting in the context of providing products to be used by others, important factors are: (A) the probability and gravity of rights infringement caused by the products provided; and (B) the ability of the product provider to recognize the occurrence of infringement. These factors are considered in a correlative manner to determine whether or not liability for aiding and abetting arises.[60]

In reference judicial precedent 6, the court found a duty to select lessees according to whether or not they had concluded or applied for a licensing agreement for the use of copyrighted works,

---

[59] For a detailed commentary, see Makiko, Takabe (2004) Hankai [Case commentaries], *Saiko Saibansho Hanrei Kaisetsu Minji-hen Heisei 13-nendo (jou) [Commentaries on supreme court judgements on civil cases in fiscal 2001 (vol.1)],* p. 196 et seq. In addition, while reference judicial precedent 6 does not directly use the term "aiding and abetting", commentaries that explain this precedent as affirming the aiding and abetting liability of the lessor include: Anonymous (2001) Hankai [Case Comment] *Hanrei Times, No. 1058,* p. 110; and Handa, Masao, et al.(eds.)(2015), *Chosakuken-ho Commentary 1 [Commentary on the Copyright vol.1] 1-jyo ~ 25-jyo [Arts. 1-25] Dai 2-han [2nd ed.] (in Japanese)*. p. 911 [Tsujita Yoshiyuki] Keiso Shobo.

[60] The judicial precedents on negligent aiding and abetting referenced in the note 17 above also identified the perspectives of (A) and (B) as important factors in finding negligence.

in light of the fact that, due to the nature of karaoke equipment, there is a high probability that leasing karaoke equipment without confirming the conclusion of or application for such an agreement will result in copyright infringement. However, image generation AI can be used lawfully for a wide variety of purposes including private creative activities. Since the probability that users will infringe rights of publicity is not typically high, and since the right of publicity is a personality right recognized by judicial precedent whose infringement is not subject to criminal penalties, among other reasons, the circumstances differ from the premises of reference judicial precedent 6 — in particular from (i), (ii), and (iv) above. Therefore, AI developers/providers should not be imposed with a high-level duty of care such as the selection of AI business users. Rather, warning to AI business users about how to use images and design measures to mitigate risks to a certain extent are considered to become an issue.

With this in mind, when evaluating (A) the probability and gravity of rights infringement caused by the AI in question and (B) the ability of the AI developer/provider to recognize the occurrence of infringement, in the context of the image generation AI service provision, the following considerations apply.

Reference judicial precedent 6: Supreme Court Judgment, March 2, 2001, *Minshu (Supreme Court Civil Case Reports), Vol. 55, No. 2,* p. 185.

- **In cases where a lessor enters into a lease agreement for karaoke equipment that is used exclusively for the purpose of screening or performing musical works for the direct viewing or listening of the public, the lessor bears a duty of care under the rules of reason to not only notify the counterparty to the lease agreement of the obligation to conclude a licensing agreement for the use of copyrighted works with the copyright holders of the musical works, but also to confirm that the counterparty has concluded or applied for such a licensing agreement with the copyright holders, before delivering the karaoke equipment.**
- **The grounds for imposing the above duty of care include the following: (i) karaoke equipment has a high probability of causing copyright infringement by operators of establishments using the equipment; (ii) copyright infringement is a criminal act that falls under penal provisions; (iii) the lessor earns business profits from leasing karaoke equipment; (iv) it is a publicly known fact that operators of establishments using karaoke equipment do not necessarily conclude licensing agreements for the use of copyrighted works at a high rate, and therefore the lessor should foresee the probability of copyright infringement unless it can be confirmed that the counterparty to the lease agreement has concluded or applied for such a licensing agreement; and (v) the lessor can easily confirm the conclusion of or application for such a licensing agreement and can take measures to avoid copyright infringement thereby.**

For example, if AI developers and providers prohibit the commercial use of generated images in their terms of use or issue warnings to prevent AI users from using generated images in ways that infringe on third-party rights, and the AI users appropriately use the AI accordingly without any publicity rights infringement occurring through the AI, then neither the seriousness nor the likelihood of rights infringement, nor the foreseeability of such infringement, can be recognized, and liability is likely to be denied. On the other hand, in situations where appropriate use is not

recognized even with explanatory measures taken by the AI developers, liability comes into consideration from the following perspectives.

Regarding (A) above, i.e., the probability and gravity of rights infringement caused by the AI in question, as described above, the probability that rights of publicity will be infringed is not typically high in the process of using an image generation AI. In addition, in order to reduce the possibility of original data being reflected in generated images and causing infringement of rights of publicity, the following measures are commonly taken in current technology: reducing the possibility of overfitting by making the data set sufficiently large; describing images in the training data with sufficiently long text (re-captioning); training in latent space; and filtering prompts to remove proper nouns for celebrities or well-known works using a large language model or similar tool (filtering) during the generation process. When there is a concern that overfitting may remain due to the small scale of the data set or for other reasons, the following additional measures may be taken: mitigating overfitting by adding regularization terms; and conducting a simple verification using image recognition AI to check whether images closely resembling real celebrities have been generated. It should be noted that not all of the above measures are required in all cases.[61] For instance, in the case of an AI model that uses a sufficiently large data set, the risk of overfitting may already be minimal even without any special additional measures. It is sufficient to take measures that are necessary and reasonable in light of the characteristics of the AI model being developed. The important thing in terms of outcome is that the probability of rights infringement is reasonably mitigated. When, as a result of such measures, the AI model has learned the abstract features common to a group of images rather than individual pieces of the original data, and the tendency to replicate specific pieces of the original data has been reduced, the probability of rights infringement described in (A) above can be assessed as low.

On the other hand, as described in 3.3.2 above, it is generally difficult for the AI business user to check each of the images output by AI against images of numerous right holders, and it is possible that the AI business user is unable to give sufficient consideration to right holders, resulting in infringement of their rights of publicity. With this in mind, if the functions of the image generation AI and the scale of the data set are such that, in response to certain specific prompts (such as instructions containing the name or main features of a celebrity), the AI model has a pronounced tendency to reproducibly generate images that closely resemble a specific celebrity, it cannot be denied that there is a possibility that the probability of rights infringement described in (A) above would be assessed as high, since it would not be appropriate to leave the prevention of rights infringement solely to the AI business user's judgment.

Regarding (B), i.e., the ability of AI developer/provider F to recognize the occurrence of infringement, since image generation AI can be used for a wide variety of purposes, F's ability to recognize the occurrence of infringement of rights of publicity is not immediately high. On the other hand, if F recognizes a sign that infringement of the right of publicity has occurred through the AI

[61] As discussed in Section 2.2.1(3) above, unlike in Hypothetical Case 4 where the AI developer and AI provider are the same entity, the content of the measures required may differ depending on their respective roles within the value chain.

that F provides (for example, if F receives a complaint from talent V), there is an increased possibility that negligence will be found with respect to any subsequent acts of providing the AI.

In case (b), the data set used contains approximately 1 billion images and is thus large-scale, and AI service developer F has taken measures such as re-captioning and training in latent space to prevent original data from being reflected in generated images without change. Additionally, no images closely resembling talent V have been generated in any cases other than the present case. For these reasons, neither (A) the probability and gravity of rights infringement caused by the tool provided nor (B) the ability to recognize the occurrence of infringement exists, and it is considered that liability for aiding and abetting does not arise.

As a note, the outline of the preventive measures described above as examples is as follows. For more detailed information on the addition of regularization terms, see "<Reference> Overview of Regularization" on p. 46 below.

Specific Measure	Timing	Details
Training in the latent space	Training/development stage	Training is conducted in the latent space that retains only the important features of the input data, in order to reduce the possibility that the AI model will memorize the original data images in pixel space and generate similar images.
Re-captioning	Training/development stage	By describing images in the original data with more generalized and abstracted expressions instead of proper nouns, the data is prevented from being associated with a specific person or the like.
Scale of data set	Training/development stage	The data set is made sufficiently large in order to reduce the possibility that the AI model will overfit the data.
Addition of regularization terms	Training/development stage	When there are concerns about overfitting, high dependence on specific images, or when a high degree of similarity is detected in similarity verification, dropout or regularization terms in the loss function with regularization effects are implemented.
Filtering (input/output data)	Generation/use stage	The risk of generating images that may infringe rights is reduced by filtering (exclusion list filter) the input and output of the AI model.

In addition to the above preventive measures regarding rights infringement, it is considered that AI service developer F can reduce the possibility of being held liable for aiding and abetting by taking measures required under the AI Guidelines for Business.[62] The measures considered

[62] As a hypothetical case, one may envisage a situation where a person whose images are included in the data set was unknown at the time the AI model was developed, but had become well known by the time the images generated by the AI were used. In this case, infringement of the right of publicity could occur in the process of using the AI, and the aiding and abetting liability of the AI developer/provider could become an issue. It is considered that, by fulfilling the rights infringement preventive

necessary may differ depending on the characteristics of F's service and individual circumstances. However, the state of governance development illustrated by the examples below would be considered circumstances that demonstrate that F exercised as much care as possible to reduce the risk of rights infringement.

<Reference> Measures Envisaged under the AI Guidelines for Business with Respect to the Right of Publicity

The table below organizes, for each party in hypothetical case 3, the measures envisaged under the AI Guidelines for Business from the perspective of reducing the risk of infringement of the right of publicity. As this table is only an overview, please refer to the relevant sections of the AI Guidelines for Business for specific details.

The measures considered necessary may differ depending on the characteristics of the AI and individual circumstances. As described above, taking the measures prescribed by the AI Guidelines for Business as necessary leads to a reduction in the possibility of being assessed as having been negligent. It should be noted that, as described in 2.1.3 above, since the AI Guidelines for Business encourages business operators to make voluntary efforts, failing to take the measures listed below does not immediately mean that there was a violation of the duty of care by each party.

			Envisaged measures regarding the right of publicity	Relevant sections of the Guidelines[63]
Apparel business operator G	AI business user	Measures for appropriate use	● Investigate and confirm within a reasonable scope whether the generated images infringe the right of publicity, and use them to an appropriate extent, etc.	U-2) i. Appropriate use with safety in mind
AI service developer F	AI developer/provider	Design measures	● Measures relating to data training ・Confirm whether images or the like of celebrities are included in the data set ・Confirm whether appropriate rights processing has been done in accordance with the provenance of the data, etc. ● Continuous monitoring ・Record logs and samples to verify whether the model reproduces the appearance of celebrities, and take improvement measures as necessary, etc.	D-2) i. Appropriate data training ※For specific examples of measures, see pages 98-99 of the Appendix D-6) i. Ensuring verifiability

measures described in the main text and other AI governance requirements, the AI developer/provider can be assessed as having taken reasonable measures to detect and prevent such potential rights infringements, thereby reducing the possibility of being found to have violated a duty of care.

[63] See the main body of the AI Guidelines for Business, pp. 29-41.

		Explanatory measures	• Provide explanations regarding the risk of infringement of the right of publicity • Establish provisions in the terms of use or other documents restricting the scope of use of generated images or prohibiting use in a manner that infringes portrait rights, in accordance with the purpose of the service, etc.	D-2) iii. Development conducive to appropriate use D-6) ii. Provision of information to relevant stakeholders P-2) ii. Provision conducive to appropriate use P-6) ii. Provision of information to relevant stakeholders

3.3.5 Case (c)

Case (c) involves a setting where AI service developer F developed an AI model that frequently generates images identical or similar to celebrities after being trained on a data set labeled with celebrities' names, and then sold the image generation AI that uses this model as an AI intended to generate various images of celebrities. In this type of exceptional case, it should be noted that the act of providing the AI itself may be assessed as an infringement of the right of publicity.

While the AI model or service for image generation itself does not constitute "images or the like," the act of providing it does not, as a rule, constitute an act of infringement of the right of publicity. However, if the act of providing the image generation AI itself is assessed as having the dominant purpose of utilizing the customer attraction of images or the like, there is a possibility that the act may be assessed as an infringement of the right of publicity as one that is substantially the same in terms of illegality as the Three Types. [64]

Under this view, for example, in cases where an image generation AI that makes it easy to generate images identical or similar to real celebrities with customer attraction is sold with this ease as a selling point, the AI has commercial value in that it allows users to freely generate images of celebrities and use them for appreciation and other purposes. This is similar to type (A) of the Three Types (i.e., using images of celebrities as goods or the like that are independent objects of appreciation, and selling them). In addition, since the ability to generate images or the like with customer attraction becomes a factor that differentiates such an image generation AI from other generative AIs, it is also similar to type (B) of the Three Types (i.e., attaching images or the like to goods or the like with the intention of differentiating the goods or the like).[65] Therefore, it is possible that the sale of an image generation AI with such features would be assessed

[64] Nakajima, the note 44 above, p. 50 notes that the "or the like" appended to the Three Types in reference judicial precedent 4 is intended to exceptionally expand the scope of the prohibitory effect of the right of publicity to cases that are assessed as substantially the same in terms of illegality as (i.e., "equivalent to") the Three Types, and that this was done with consideration for the fact that unpredictable manners of use may arise in the future.

[65] With respect to type (B) of the Three Types, Nakajima, the note 44 above, pp. 47-49 explains that it is a type that makes unlawful so-called "character goods" that use images or the like. The author further explains the difference between types (A) and (B) as follows: type (A) focuses on the value of appreciation of "the image itself" and commercializes this, whereas type (B) focuses on differentiating the "product as a whole" through the use of images and commercializes this as character goods.

as falling under "or the like" in the Three Types as an act that is substantially the same in terms of illegality as those in the Three Types and that has the dominant purpose of utilizing the customer attraction of images or the like.[66] In such case, it is considered that, regardless of the act of the AI business user, a tort liability based on infringement of the right of publicity is established with respect to the act of providing the AI itself.[67]

<Reference> Overview of Regularization

The AI mainly discussed in this Guidance is one that extracts statistical trends and patterns behind data by a method called "machine learning" and utilizes them for tasks such as recognition, prediction, and generation. However, when an AI is trained, if it becomes excessively adapted to the data, which are in fact only samples, the noise contained in the data may be reflected, causing the resulting pattern to lose generality (overfitting).

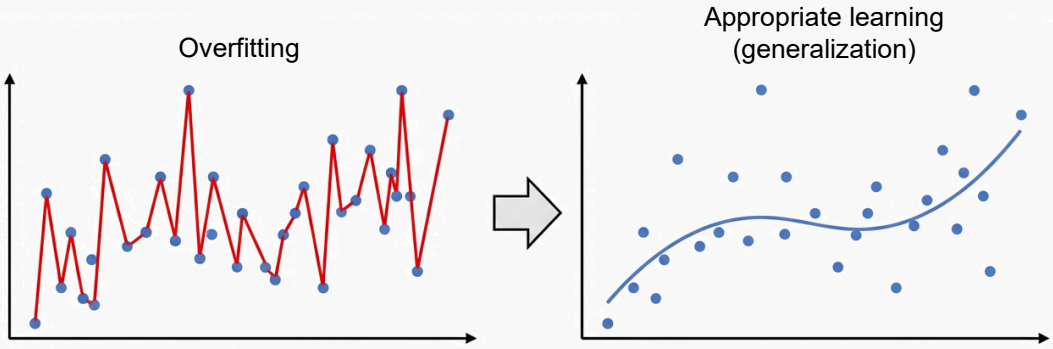


Figure: Conceptual diagram of overfitting and appropriate learning

When overfitting occurs, rather than extracting general patterns from training data, the AI tends to behave in a manner that is overly dependent on the features of individual data items. As a result, there is a risk that an image generation AI may output images similar to training data, or that a transaction screening AI may fail to perform sufficiently for unknown cases, which may be one factor that leads to some individuals with certain attributes being evaluated unfavorably. In order to prevent such risks, there is a technique called "**regularization**" that prevents excessive dependence on specific data.

However, since regularization mitigates overfitting by controlling the complexity of the model, it does not correct biases that originate from the composition of the training data or the setting of the target variable itself. Therefore, it needs to be used in combination with other measures.

In machine learning, a performance indicator called a "**loss function**"※1 is set initially, and the model is trained by repeatedly updating the parameters to minimize it (in the example shown in the figure above, the relevant indicator is one calculated from the distance between the model's predicted values and the actual data). In doing

[66] In cases where an image generation AI like the one in case (c) is sold, it is also foreseeable that talent V's name, images, or the like may be used in the product name, packaging, sales web pages, and other materials in the course of advertising and other activities. In such cases, the use would fall under type (B) of the Three Types, i.e., attaching images or the like to goods or the like with the intention of differentiating the goods or the like, and/or type (C), i.e., using images or the like in ads for goods or the like. In this respect as well, tort liability based on infringement of the right of publicity may arise.

[67] There is also a possible view against the approach presented in the main text, which emphasizes the fact that AI service developer F is merely providing an image generation AI and that talent V's images are generated and used only in the hands of AI business users, and takes the position that F's act does not fall within "or the like" and therefore cannot constitute an infringement of the right of publicity.

so, if the sole aim is to minimize the prediction error, the model tends to become an extreme model like the one shown on the left side of the figure above. For this reason, rather than looking at only the prediction error as the loss function, a term representing the complexity of the model※2 (a regularization term) is incorporated as follows:

Loss = Prediction error + λ × Complexity

λ is a coefficient (hyperparameter) that determines the strength of regularization. By mitigating excessive reactions to slight differences in input or to specific data, it guides the model to extract patterns common to the data as a whole, thereby helping to suppress the phenomenon of the model memorizing the training data verbatim and outputting it. Increasing λ to strengthen regularization may reduce the risks arising from "memorization" of training data, while there is a trade-off in that the quality of the model's output tends to be more easily affected.

※1 The loss function can be intuitively understood as a kind of scorecard that the AI refers to during training. It is a statistical numerical expression of, among other things, how much deviation occurs between predicted values and actual measured values. If the value can be reduced, the model is assessed as having become more "superior."

※2 The "complexity" referred to here is technically defined by multiple indicators such as the sum of squares of the model's weights, but can be intuitively understood as an indicator designed to guide the model so that it does not learn in a manner that is overly dependent on specific data.

3.4 Hypothetical case 4: Transaction Screening AI

AI service developer H has developed and provided an AI for residential tenancy screening that calculates credit scores for applicants for rental properties. The AI service is composed of a machine learning algorithm; when real estate agent users input prescribed information relating to applicants, the system quantitatively calculates a credit score based on that information, classifies applicants as approved or rejected based on a certain threshold, and presents the score and the approval/rejection result to the user. I, a real estate rental business operating nationwide, adopted H's AI service and was streamlining and reducing the staffing of its screening operations for its own properties.[68]

[Case (a)] In the above AI, no training was performed that focused on constitutionally prohibited discrimination attributes; however, as a result of training that placed excessive weight on information relating to income and employment type, statistical bias that tended to treat women

[68] As a separate example from hypothetical case 4, one might consider a case where, in a real estate brokerage business, a real estate business operator uses transaction screening AI to conduct risk assessment and a property owner makes the ultimate decision based on the assessment results; however, in this hypothetical case, the introduction of AI for properties owned by the business operator itself is assumed in order to avoid complicating the facts of the case.

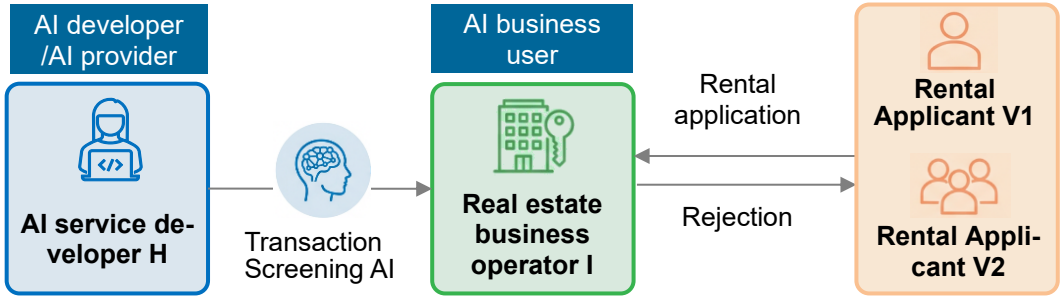
significantly less favorably than men was included. H was aware of this bias but took no particular corrective measures and provided the service to users without explaining the content or risks of the bias. The statistical tendency to treat women significantly less favorably than men was an evaluative tendency that was not particularly necessary for the purposes of tenancy screening.

On one occasion, when female applicant V1 applied, the AI rendered a rejection decision, and I's person in charge I' relied on the AI's judgment and rejected the application. At that time, I' did not specifically recognize the existence of the above bias, but had a vague awareness that, after the introduction of the AI, the rate at which female applicants were rejected had increased unnaturally compared to before the introduction.

[Case (b)] H took measures to give due consideration to bias during the development stage, avoiding training the AI model on attribute information such as religion or personal beliefs, and setting only evaluation factors that had no apparent connection to religion or personal beliefs, such as address, occupation, years of service, annual income, rent-to-income ratio, and payment delinquency history. As a result, H confirmed in the fairness assessment conducted at the time of AI model development that there were no concerns about bias that could lead to discrimination. I adopted the AI on this basis, implemented data management to avoid inputting data that could give rise to bias, and requested H to periodically monitor for the occurrence of bias during operations, while using the AI for screening.

After I adopted the above AI, a large community V2 was formed that shared political beliefs regarding work styles and settled in the same area. Since members of V2 frequently changed jobs over short periods, the above AI, through its continuous learning process, learned the tendency of short-term workers to be concentrated in that area, and came to treat residence in that area itself as one risk factor. As a result, members of V2 who resided in Area X and tended to change jobs frequently were rejected at a notably high rate, and multiple members received rejection decisions and had their applications turned down.

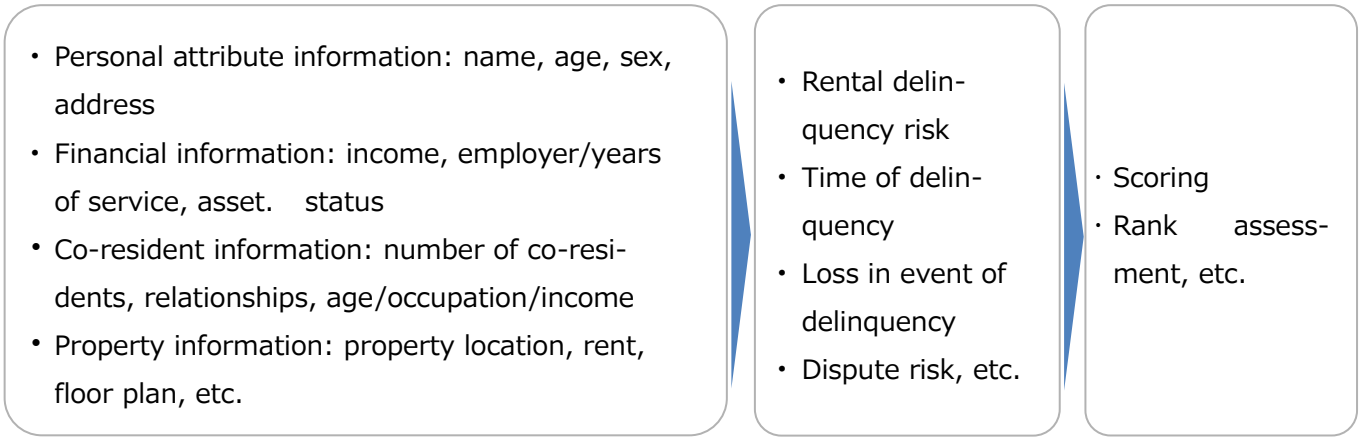
H and I were completely unaware of the above bias until a certain point in time; however, upon detecting the above bias through a jointly conducted monitoring process, they took measures such as, for the time being, having human reviewers rather than the AI conduct screening for applicants in that area.



上 In the above cases, if the AI's bias is assessed as having led to a refusal to transact that lacks social acceptability, the issue is whether or not H and I are liable to V and others (Article 709 of the Civil Code).

3.4.1 Transaction Screening AI and Fairness

The use of AI in settings such as transaction screening has been growing in recent years. One type that has been commonly used is the type that scores the degree of transactionally significant risk by taking into account various attributes of the person being evaluated (see the figure below) [69]. The use of AI is expected to enable highly accurate risk analysis that takes into account a wide range of factors, and to reduce to a certain degree the human-caused bias in judgment arising from cognitive biases on the part of the person in charge, among other things.



入 Tenancy screening, like transaction screening in general, is governed by the principle of freedom of contract, and because there is discretion in the selection of counterparties and in the setting of transaction conditions, such decisions are in principle not assessed as unlawful.

On the other hand, in cases assessed as a refusal to transact that lacks social acceptability on grounds of attributes that may lead to discrimination (hereinafter "sensitive attributes")[70], such as attributes enumerated in Article 14, paragraph (1) of the Constitution, there may be a possibility that such refusal constitutes a tort based on infringement of moral rights[71] (see reference judicial

[69] In recent years, research and implementations that use large language models in transaction screening, while aiming for fairness through prompts such as "use no bias," have been advancing (cf. Bowen III, D., Price, S. M., & Yang, K. (August 20, 2025). *AI exhibits racial bias in mortgage underwriting decisions.* Lehigh University News. https://news.lehigh.edu/ai-exhibits-racial-bias-in-mortgage-underwriting-decisions, etc.); however, hypothetical case 4 assumes a conventional AI that performs scoring.

[70] The extent to which fairness is required among individual characteristics varies depending on the cultural and social background of each country or region, and it is difficult to determine definitively. Therefore, in hypothetical case 4, the focus is on a refusal to transact based on sex or personal beliefs, which are at minimum prohibited from discrimination under Article 14, paragraph (1) of the Constitution, and the points of focus in the discussion of liability are examined.

[71] There are also cases where the act is found to constitute a tort based on a breach of an obligation arising from the duty of good faith incidental to the transaction (cf. Kyoto District Court Judgment, October 2, 2007 (unreported), etc.).

precedents 7 through 11 below). In settings where AI is used, the issue of when such conduct constitutes a tort is a legal issue.[72]

3.4.2 The Position of AI in Hypothetical Case 4

Transaction screening AI can effectively evaluate credit risk and the like by taking into account various attributes of the person being evaluated, and functionally can substitute for the judgment of AI business users in transaction screening. On the other hand, if the AI's output contains a certain degree of bias, there is a risk that this will lead to a refusal to transact that lacks social acceptability; and because it is considered that human assessment and verification during the course of use are required to prevent such output from being relied upon, transaction screening AI also falls under assistive/supportive AI.

If the output of an assistive/supportive AI includes something that infringes third-party rights, the AI business user is required to use the AI's output to an appropriate extent at the user's own judgment. In particular, in cases such as hypothetical case 4, where AI output is used in making the ultimate judgment of whether or not to transact, it should be noted that there may be a possibility that directly relying on the AI's output for such judgment could be assessed as a refusal to transact that lacks social acceptability. However, as stated above, because in many cases it is technically difficult to assess whether or not there is a problem with individual AI output from the perspective of bias, it is considered that measures such as collecting information about the AI model's bias would be at the center of the duty of care. In addition, it is an issue whether or not the AI developer/provider should have taken certain design and explanatory measures in order to enable the AI business user to use the AI's output appropriately (see 3.4.4 and below).

3.4.3 Requirements for a Refusal to Transact to Constitute a Tort

In conventional judicial precedents, the fact that a person refused to transact after focusing on sensitive attributes has provided a common basis for both the rights infringement requirement and the intent/negligence requirement, and many decisions have not distinguished between the two separately.

On the other hand, in cases involving the use of AI, as described below, the question of rights infringement — i.e., whether the AI's output legally constitutes an infringement of moral rights — and the question of intent/negligence — i.e., whether or not each party bore a duty to prevent

---

[72] The report of the Research Group on Legal Issues Surrounding the Use of AI by Financial Institutions established by the Institute for Monetary and Economic Studies, Bank of Japan (Institute for Monetary and Economic Studies, Bank of Japan (2025). *"Kinyukikan ni okeru AI no Rikatuyo wo Meguru Houritsumondai Kenkyukai" Houkokusho [Report of the "Research Group on Legal Issues Surrounding the Use of AI by Financial Institutions" — Private Law Risks and Management Associated with the Use of AI by Financial Institutions]* p. 21) also states that, for example, in cases where unfair bias is reflected in the AI's judgment in loan screening using dependable/substitutive AI, this could constitute a violation of the obligation in good faith to conduct loan screening fairly.

such rights infringement — may differ qualitatively. Therefore, these are organized separately below.

(1) Rights Infringement

In the conventional setting where a person conducts transaction screening, judicial precedents have determined rights infringement primarily based on (C) the rationality of the purpose and (D) the appropriateness of the means, taking into account (A) the degree of importance of the counterparty's attribute in that transaction and (B) the degree of disadvantage to the counterparty, on the premise that it was possible to establish as fact what factors were primarily considered in the ultimate judgment (whether it was a sensitive attribute or other factors).

In other words, it is not the use of sensitive attributes in transaction screening itself that is problematic; rather, there may be an assessment of unlawfulness only in exceptional cases where in particular the rationality of the purpose or the appropriateness of the means is not found.

It is considered that the same criteria apply when screening is conducted by AI; however, because the basis for an AI's judgment is often not clear from a mere glance at the output score or assessment result, in many cases it is technically difficult to establish as fact whether or not a sensitive attribute was given primary weight in individual screenings.

Therefore, the starting point for determining the presence or absence of rights infringement is considered to be verifying, by means of fairness assessment indicators such as those shown in the table below, the degree to which evaluation tendencies differ between subjects with and without sensitive attributes (i.e., whether sensitive attributes are the primary factor in screening).

Indicator	Content	Example
Equal Opportunity	Whether the rate at which AI correctly identifies persons who should have been approved (with a correct label of 1) as approved (True Positive Rate, "TPR") is the same regardless of attribute	If TPR for men = 80% and TPR for women = 60%, 20% of women who should be approved are being rejected
Calibration Difference	Whether the predicted scores assigned by the AI match, for each attribute group, the actual probability of approval	If the AI predicts a score of 0.8 for a man, 80% of men with that score actually pass; but if the AI predicts a score of 0.8 for a woman, only 60% actually pass
Demographic Parity[73]	Whether the rate at which AI approves applicants is the same for each attribute group	The AI approves 60% of men and 40% of women, a 20 percentage point gap

[73] Where there are actual differences in ability levels — such as income — among groups of applicants, requiring equal pass rates may create reverse discrimination or operational disadvantages; therefore, caution should be exercised regarding the use of this indicator.

If the results of such fairness assessment indicate that sensitive attributes are the primary factor in screening, assessments (A) through (D) below are conducted. With specific reference to the tenancy screening in hypothetical case 4, the analysis is as follows.

(A) The degree of importance of the counterparty's attribute in that transaction

- ✓ In the setting of concluding a lease agreement that continues over a certain extended period, given that in general it is difficult for the lessor to terminate the contract mid-term, the attributes of the lessee entering into such a long-term contractual relationship are generally an important consideration factor for the lessor, and the lessor's interest in taking the counterparty's attributes into account in screening is respected.

(B) The degree of disadvantage to the counterparty

- ✓ Even if the person being evaluated is refused a transaction, it is considered possible for the person to conclude a contract for another rental property; however, contracting for a rental property is an important opportunity that forms the foundation of one's living.
- ✓ Furthermore, if a person being evaluated receives a low rating from a general-purpose AI service used by multiple lessors, the person would have fewer properties to choose from, resulting in a greater disadvantage to that person; on the other hand, if a person being evaluated receives a low rating from a service developed exclusively for a specific lessor, the person can choose a different property not subject to that AI's screening, so the disadvantage is relatively smaller. Nevertheless, the above interest of the person being evaluated in building the foundation of their living should be given sufficient consideration.

(C) The rationality of the purpose

- ✓ In the context of the business purpose, there must be a rational reason for the tendency to evaluate persons with sensitive attributes unfavorably.[74] With regard to tenancy screening, what is important is the connection between such attributes and the risks associated with tenants (such as rent delinquency or residential disputes, among other things).
- ✓ In establishing the rationality of the purpose for screening using AI, it is considered important, among other things, to select attributes whose consideration is rationally grounded when determining the attributes to be trained on, and to examine alternative evaluation factors in cases where a rational basis cannot necessarily be established.[75]

(D) The appropriateness of the means and availability of alternatives

---

[74] For example, there is a case where the rationality of the purpose was found with respect to a provision requiring permanent residency status for mortgages for foreign nationals, taking into account, among other factors, the high necessity of having standardized conditions in connection with business purposes such as ensuring repayment capacity (see reference judicial precedent 8 below).

[75] See the AI Guidelines for Business Appendix, p. 98.

- ✓ In the context of the purpose, it would be questioned whether it is rational to uniformly refuse to transact with persons who have sensitive attributes, whether consideration was given to examining alternative means of conducting screening by focusing on more indirect and specific attributes, and whether or not such alternative means exist.
- ✓ With regard to screening using AI, if necessary measures such as those envisaged under the AI Guidelines for Business as listed below[76] are taken, it is considered that this would weigh in favor of establishing the appropriateness of means:
  ・Implementing mitigation measures through review of data ratios and labels (a certain degree of bias still remains despite such measures);
  ・Adopting other consideration factors to reduce bias toward sensitive attributes;
  ・Implementing regularization with a fairness penalty term;[77]
  ・Designing the scope requiring human involvement to focus on the range where bias is likely to affect approval/rejection outcomes; and the like.
- ✓ Furthermore, if a procedure is established for objecting to screening results for persons who dispute the result, it is considered that this would be weighed as a measure that mitigates the disadvantage to persons being evaluated and would contribute to enhancing the appropriateness of means.

Although the determination of rights infringement requires a comprehensive consideration of individual cases, tendencies can be summarized as follows.

- ✓ In cases where the assessment results indicate a tendency for sensitive attributes to be the primary factor in the statistical processing conducted by the AI, and where it is not possible to explain that such evaluation tendency is rational in light of the purpose of the screening, or where measures to mitigate the disadvantage to the subject have not been taken, this will weigh in favor of establishing rights infringement.
- ✓ Even where sensitive attributes are the primary factor, if this is rational in light of the purpose of the screening and mitigation measures for the disadvantage to the subject have been taken where necessary, there will not necessarily be a finding of rights infringement.
- ✓ In cases where an evaluation algorithm has been adopted that is centered on evaluation factors that are neutral with respect to sensitive attributes and that are commonly used in screening (such as employment type, occupation, and income), it will be easier to find that the rationality of the purpose and the appropriateness of the means are established. However, it should be noted that such neutral factors may function as proxy variables, resulting in a tendency to unfavorably evaluate persons with sensitive attributes without rational justification.

---

[76] See "<Reference> Measures Under the AI Guidelines for Business Regarding Fairness" on p. 59 below.
[77] See "<Reference> Overview of Regularization" on p. 46 above.

Reference judicial precedent 7: Provision of women-only train cars (Tokyo District Court Judgment, July 12, 2011 (unreported))

- **The provision of women-only train cars can be construed as being aimed at providing female passengers, who are at risk of falling victim to groping, with a more comfortable and reassuring environment for commuting to work or school, in trains in metropolitan areas such as the Tokyo area that are significantly crowded during rush hours on weekdays. This is considered to be legitimate in terms of purpose.**
- **Women-only train cars are set in only some trains during part of the rush-hour period on weekdays, and amount to only one car out of a six-car train; this is not something that makes it difficult for able-bodied adult male passengers to use the other cars to reach their destination, and does not impose any particular disadvantage on able-bodied adult male passengers.**

  → The court found no tortious liability on the grounds of the rationality of the purpose and the appropriateness of the means (in particular, that the cars were operated in a limited number and only during limited hours).

Reference judicial precedent 8: Permanent residency requirement in residential mortgage screening (Tokyo District Court Judgment, November 12, 2001, *Hanrei Jiho, No. 1789,* p. 96; Tokyo High Court Judgment, August 29, 2002, Kinyu Shoji Hanrei, No. 1155, p. 20)

- **The act of a bank rejecting a mortgage application from a foreign national without permanent residency status did not violate Article 14 of the Constitution or other provisions.**
- **Since residential mortgages involve high costs associated with the loan (such as contract processing fees, debt/collateral management costs, and debt collection costs) due to their long repayment periods and other factors, and cannot be said to be highly profitable products, banks need to handle a large volume of mortgages to achieve adequate profitability, and for this purpose it is necessary to standardize the requirements for eligible borrowers as much as possible.**
- **Taking into account, among other things, that foreign nationals without permanent residency have their residence period restricted to three years, and that if they leave Japan the costs, time, and effort required for debt management and debt collection would become enormous, it is not irrational to make a standardized determination of eligibility for financing based on the presence or absence of residence status.**

  → In a transaction where the counterparty's attributes are important, such as a mortgage, taking into account the bank's freedom to conduct business and the need to have standardized transaction conditions, and considering the increase in debt collection costs resulting from the borrower leaving the country, the court found rationality in uniformly requiring permanent residency status.

Reference judicial precedent 9: Refusal of bathing service at a public bath (Sapporo District Court Judgment, November 11, 2002, *Hanrei Jiho, No. 1806,* p. 84)

- **Taking into account the public nature of the facility, which was licensed by the governor under the Public Bath Act, the court held that uniformly refusing access to all foreign nationals was clearly irrational, and that the facility should have individually instructed or dealt with persons who caused inconvenience to other bathers or who did not follow bathing etiquette.**

→ **Taking into account the public nature of the defendant's facility, the court found the tortious liability indicating that the facility should have taken alternative measures such as individually instructing and dealing with patrons, rather than uniformly refusing entry to persons based on nationality.**

Reference judicial precedent 10: Refusal of membership in a golf club (1) (Hamamatsu Branch of Shizuoka District Court Judgment, September 8, 2014, *Hanrei Jiho, No. 2243,* p. 67)

- **In a case involving the refusal of membership to the plaintiff, who had received a family court ruling changing their legal sex to female, the court held that the defendant golf club was not recognized as a closed organization, among other reasons, because no strict membership requirements existed; and that while the strong anxieties and feelings of discomfort among existing members concerning the plaintiff using the defendant's golf course were nothing more than abstract concerns, the emotional distress to the plaintiff was not negligible. Therefore, the defendant's refusal of membership was unlawful.**
  → **The organization was not found to be a closed entity, and the tort liability was recognized after taking into account the rationality of considering the sex change and the degree of disadvantage to the plaintiff.**

Reference judicial precedent 11: Refusal of membership in a golf club (2) (Tokyo High Court Judgment, January 23, 2002, *Hanrei Jiho, No. 1773,* p. 34; original judgment: Tokyo District Court Judgment, May 31, 2001, *Kinyu Hanrei, No. 1138,* p. 35)

- **In a case where the golf club was found to have a highly closed nature in light of the strictness of its membership procedures and other factors, the establishment of a tort was denied even though the club uniformly restricted membership for foreign nationals.**
- **As factors indicating the small degree of disadvantage to the plaintiff, it was pointed out that, even if membership status had been obtained, the only benefit would have been convenience in using the golf course and the like.**
  → **Unlike reference judicial precedent 10, the organization was found to be a highly-closed entity and the court found no tortious liability even in a case involving a uniform nationality requirement.**

(2) Intent/Negligence

If rights infringement is recognized in (1) above, the issue is what duty of care each party bears.

As stated above, because it is technically difficult in many cases to assess whether or not there is a problem with individual AI output from the perspective of bias, the duty of care of AI business users is considered to be centered on, within a reasonably possible scope, confirming matters related to the overall fairness of the AI model, and, depending on the associated risk, involving human judgment or examining the appropriateness of introducing the AI, and the like. In addition, this is a category in which the AI developer/provider is also required to take certain design and explanatory measures from the perspective of enabling the AI business user to use the AI's output appropriately.

3.4.4 Case (a)

Case (a) is a case in which a bias concern existed at the time of development and yet the AI was used for screening, leading to a refusal to transact. In such a case, each party may bear the following liability.

(1) Rights Infringement

In case (a), the following are factors that increase the possibility of being assessed as unlawful rights infringement: the fact that I operates a real estate rental business nationwide and the scope of the AI's impact is large; the fact that the AI has an evaluation tendency that tends to treat women significantly less favorably than men; the fact that H took no particular corrective measures with respect to such tendency; and the fact that such tendency was an evaluative tendency not particularly necessary for the purposes of screening.

Although the AI does not directly evaluate sensitive attributes unfavorably, it should be noted that, where as a result sensitive attributes have a large influence on screening outcomes and it is not possible to explain that such evaluation tendency is rational in light of the purpose of the screening, or where measures to mitigate the disadvantage to the subject have not been taken, there is a possibility that this could be assessed as rights infringement.

If relying on the AI's output to make a refusal to transact is assessed as rights infringement, the issue is what measures H and I should have taken to prevent the rights infringement.

(2) Liability of I as AI Business User

Because the AI in hypothetical case 4 is an assistive/supportive AI, I bears a duty to conduct screening fairly in relation to applicants regardless of whether or not it uses AI for screening; and if it uses AI, it must not directly rely on the AI's output — if it infringes rights — as the judgment in the screening. However, as stated above, given that in many cases it is difficult to detect whether there is a problem from the perspective of bias by merely looking at individual scores or assessment results output by the AI, it is necessary to take into account factors other than the output in order to use the AI's output to an appropriate extent in making the ultimate screening judgment. Therefore, the duty of care borne by I is considered to be centered on, within a reasonably possible scope, confirming matters related to the overall fairness of the AI model rather than assessment and verification of individual output, and, depending on the associated risk, involving human judgment or examining the appropriateness of introducing the AI, and the like. In this regard, the AI Guidelines for Business require AI business users to take measures such as collecting information about AI model bias as necessary, and having human judgment involved while giving due consideration to bias (see "<Reference> Measures Envisaged under the AI Guidelines for Business with Respect to the Fairness" on p. 59 below,).

In case (a), taking into account the fact that there was a notable bias concerning sex at the time I introduced the AI for screening, and that I's person in charge I' had a vague awareness that the rate of rejection of female applicants had unnaturally increased after the introduction of the AI, there is a possibility that I could have recognized the risk of rights infringement and avoided it through simple measures such as collecting basic information about the AI model's bias. Therefore, I may be found to have violated its duty of care for failing to collect information from H about the fairness of the AI model, request H to take bias reduction measures where necessary, consider adopting an AI with less concern about bias, or take measures such as having human judgment involved while using the AI's output as a reference only.

(3) Liability of H as AI Developer/Provider

As a provider of an assistive/supportive AI, H is subject to scrutiny as to whether, as described in 2.2.1 above, (A) it provided a clear explanation of the AI's functions, performance limitations, methods of use, significant risks, and the like, and whether (B) it should have taken certain measures to prevent rights infringement with respect to risks that the AI business user could not necessarily have easily foreseen or addressed. As stated above, given that in many cases it is difficult for the AI business user to detect whether there is a problem with individual output from the perspective of bias, H is considered to be required either to take mitigation measures and the like within a reasonable scope with respect to bias risks of which it is aware, or to clearly explain the residual risks to I.

In case (a), given that there was a notable bias concerning sex at the time of AI development and yet H did not explain this point to the AI business user, it is highly possible that H would be assessed as having failed to take the required measures from either perspective (A) or (B). Specifically, H may be found to have violated its duty of care for: failing to take measures to prevent unnecessary bias from affecting screening results, such as by reviewing the weighting of evaluation factors; failing to explain an overview of the bias to I; and failing to recommend the involvement of human judgment where necessary, among other things.

3.4.5 Case (b)

In contrast to a case such as case (a), there remains a possibility that unexpected bias will arise even where the parties have taken careful measures to address bias during development, provision, and use. In case (b), the issue is the determination of liability with respect to bias that arose as a result of changes in the external social structure.

(1) Rights Infringement

In case (b), whether or not the significant evaluation tendency to reject persons who belong to V2 constitutes rights infringement depends on the factors behind such tendency. If this evaluation tendency can be explained by factors commonly considered in screening (such as employment

type, occupation, and income), there is a higher possibility that the rationality of the purpose and the appropriateness of the means are found, and the tendency would not be assessed as rights infringement. On the other hand, if a factor with low necessity in screening has influence, and circumstances establishing the rationality and appropriateness of adopting such factor are scarce, yet the tendency to unfavorably evaluate persons who belong to V2 has become notably pronounced, there is a risk that this could be assessed as rights infringement.

Even if rights infringement is found, the issue is whether or not H or I can be found to have been negligent, in light of the process by which the bias arose as described above.

(2) Liability of I as AI Business User

As in 3.4.4(2) above, I may bear a duty of care to confirm, within a reasonably possible scope, matters related to the risk of bias, and, depending on the associated risk, to involve human judgment or to examine the appropriateness of introducing the AI, and the like. However, since the bias in case (b) arose as a result of changes in the social structure after the introduction of the AI, it is considered to have been extremely difficult to foresee the risk of bias arising toward V2 and to examine the required measures; therefore, at least at the time of introduction of the AI, I cannot be found to have been negligent.

On the other hand, in cases where biases such as those in Case (b) emerge retrospectively and reach a level that can be regarded as an infringement of third-party rights, if reasonably feasible measures to prevent the direct use of AI outputs have not been taken, negligence may be recognized in that regard. As part of such measures, With regard to potential bias such as that involved in case (b), the AI Guidelines for Business require collaborative measures to detect potential bias, such as the AI business user providing feedback to the AI provider and others within a reasonable scope regarding output results, while the AI provider periodically or from time to time monitors bias trends as necessary (see "<Reference> Measures Envisaged under the AI Guidelines for Business with Respect to Fairness" on p. 59 below). If I took such measures in coordination with H, and if the risk was nonetheless assessed as not being reasonably foreseeable, negligence would not be found.

On the other hand, there is a possibility that negligence may be found from the point in time when the onset of the above bias became objectively recognizable — for example, when signs of bias became apparent through monitoring by I.[78]

(3) Liability of H as AI Developer/Provider

In light of the process by which the bias arose in case (b), it is considered to have been extremely difficult, at least at the stage of development or at the time of commencement of service

[78] In hypothetical case 4, H and I took measures to avoid rights infringement at the point in time when they detected signs of bias; however, in cases where, despite recognizing the possibility of rights infringement, reasonable measures were not taken, such failure may be assessed as negligence through omission.

provision, to foresee the risk of bias arising toward V2 and to examine mitigation measures and the like; therefore, negligence is not found in this respect. On the other hand, the AI Guidelines for Business require a periodic evaluation of inputs/outputs and the basis for judgments based on feedback from AI business users and others, as well as ongoing monitoring of evaluation tendencies; the issue is whether it was possible to examine necessary designs and explanations after taking such measures. If it was assessed that the risk was not reasonably foreseeable despite taking such measures, no negligence in design or explanation would be found.

On the other hand, there is a possibility that negligence may be found from the point in time when the above bias became objectively recognizable through monitoring, or in cases where monitoring and other measures were not taken at all. Given that H, as the developer of the algorithm, is in a position with technical and practical expertise regarding bias assessment and analysis, the point in time at which the existence of bias is considered reasonably detectable may be earlier for H than for the AI business user.

<Reference> Measures Envisaged under the AI Guidelines for Business with Respect to Fairness

Organizing the measures contemplated under the AI Guidelines for Business regarding fairness by party in hypothetical case 4 yields the following table. Since the table is only a summary, please refer to the respective sections of the AI Guidelines for Business for specific details.

Although the measures considered required vary depending on the characteristics of the AI and individual circumstances, as described in 3.4.3 and below, taking the measures provided for in the AI Guidelines for Business as necessary may contribute to reducing the possibility of being assessed as having committed rights infringement or negligence. Note that, as described in 2.1.3 above, given the nature of the AI Guidelines for Business as promoting voluntary efforts by individual businesses, a failure to take the measures listed below would not directly result in a finding of a violation of the duty of care by each party.

		Measures contemplated regarding fairness	Relevant section in the Guidelines
Real estate business operator I	**AI business user**	● Consideration given to bias contained in prompts ● **Collecting information about AI model bias** as necessary ● **Involving human judgment** while giving due consideration to bias ● Providing output results to relevant parties within a reasonable scope	U-3) i. Consideration given to bias contained in input data or prompts (Main text, pp. 40; separate volume, p. 168-169)

AI service developer H	AI provider	● Examining bias in data fairness, referenced information, and linked external services, etc. ● **Periodically evaluating** AI model **inputs/outputs and the basis for judgments** ● **Monitoring** for the occurrence of bias and examining necessary measures ● Examining the impact of bias on AI business users or non-business users through **UI** and **business process design**	P-3) i. Consideration given to bias in the configuration and data of AI systems/services (Main text, pp. 37–38; separate volume, pp. 140–142)
	AI developer	● **Determining sensitive attributes** in consideration of Article 14, paragraph (1) of the Constitution and the like ● Mitigation measures throughout pre-training, during training, and post-training with respect to **bias contained in data** (ensuring representativeness, detecting and removing latent bias, etc.) ● **Assessment of algorithm fairness** using fairness indicators (metrics) ● Designing the scope in which **human judgment** should be made ● Ensuring transparency to relevant stakeholders regarding model characteristics, etc.	D-3) i. Consideration given to bias contained in data (Main text, p. 29; separate volume, pp. 98–99) D-3) ii. Consideration given to bias contained in AI model algorithms, etc. (Main text, p. 30; separate volume, pp. 106–107) D-6) ii. Providing information to relevant stakeholders (Main text, pp. 31; separate volume, pp. 115-116)

Chapter 4 Cases Potentially Involving Dependable/Substitutive AI

This chapter examines cases that may potentially involve dependable/substitutive AI. Examples of systems that may qualify as dependable/substitutive AI include high-precision visual inspection AI used as software, as in Hypothetical Case 5 (see 4.1 below), and autonomously navigating robots combining software and hardware, as in Hypothetical Case 6 (see 4.2 below).

In addition, systems such as AI agents may also qualify as dependable/substitutive AI in the future. See 4.3 below — Hypothetical Case 7 presents, as a supplementary discussion, a scenario involving an AI agent deployed for corporate customer support, and examines the basic perspectives applicable when such a system falls within either the dependable/substitutive AI category or the assistive/supportive AI category.

In assessing whether a system qualifies as dependable/substitutive AI and in examining the specific content of each party's liability, the role and precision of the AI in light of the nature, purpose, and risks of the task in question are considered important factors.

As in Hypothetical Case 5, where the scope of work is limited to contracted inspection services, it is considered relatively straightforward to classify an AI as dependable/substitutive AI when it performs the relevant task with high precision. On the other hand, as in Hypothetical Case 6 involving warehouse transport operations, where significant risks such as collisions with workers are present, a comparatively high level of safety is required. It is considered that, in entrusting part of the transport operations to the autonomous navigation function of a robot, this falls within the category in which it is necessary to establish the appropriate environment for safe navigation and to secure the requisite level of safety by restricting the operational range to specific environments and conditions. Furthermore, even in settings such as Hypothetical Case 7, where important information bearing on consumers' purchasing decisions is provided, a comparatively high duty of care tends to be recognized from the standpoint of consumer protection, and accordingly the standards required of dependable/substitutive AI are also elevated.

Against the background of these structural differences, even within the dependable/substitutive AI category, the specific measures that AI business users are required to take, and the specific content of the liability of AI developers and manufacturers who sell or provide AI systems or robots, differ accordingly. Moreover, in relation to robots such as those in Hypothetical Case 6, it is also necessary to examine the manufacturer's liability in light of the product liability framework.

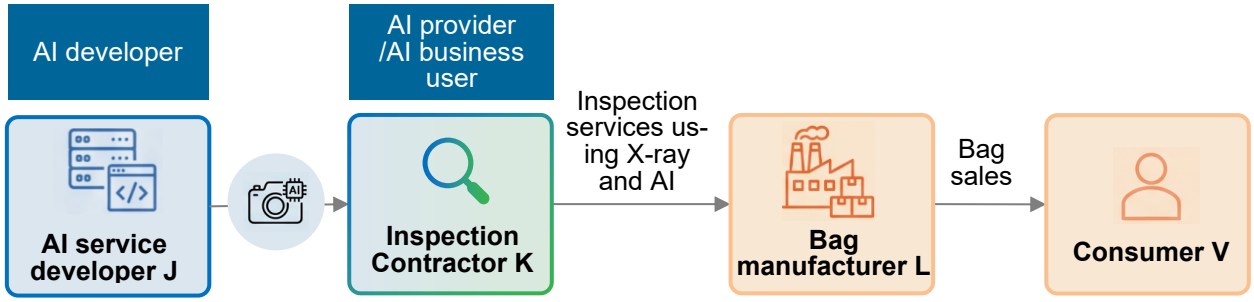
4.1 Hypothetical case 5: Visual Inspection AI

Company K[79], which provides contracted inspection services for manufactured goods, combines an image recognition AI developed by AI service developer J with an X-ray inspection device to offer an inspection service capable of detecting foreign objects such as metal fragments in manufactured goods with high precision.

Bag manufacturer L regularly used K's inspection service for the detection of metal fragments during 100% inspection. In relation to the contracted work from L, K adjusted the AI so that, under specific conditions, the detection probability for foreign objects of 1 mm or larger was approximately 98%. The AI was also used for primary screening: by directing images with high uncertainty under AI judgment to human secondary review, the service was able to process inspections with greater precision than conventional manual-only inspection in approximately half the time.

During the first three years following commencement of the contract, no serious oversights occurred in K's inspections, and only a few complaints arose regarding contamination by paper or fabric scraps. However, in the fourth year after commencement of the contract, K failed to detect a box cutter blade (2 cm) that had been mixed into a bag, and consumer V, who had purchased the bag from L, was injured. Investigation based on lot numbers and AI logs revealed that the blade was a foreign object within the AI's detection range, but it had been concealed behind a component of the bag, and the AI had missed it at the screening stage with an extremely low probability.

The business consignment agreement between K and L contained only an abstract provision that "K shall perform the inspection work in accordance with L's instructions with commercially reasonable care and skill," and no agreement had been reached regarding the specific standard of the service.



[79] This hypothetical case assumes a scenario in which K's quality management division introduces the AI system developed by J, integrates it into K's own X-ray inspection system, and places it in service for the operations of K's inspection division. In this case, K holds the status of both an AI provider and an AI business user (see the AI Guidelines for Business, main text, p. 5; Annex, pp. 9). Although K is not in a position to provide AI externally, it should be noted that under the AI Business Operator Guidelines, an AI provider is defined as “business operators who incorporate AI systems into applications, products, or existing systems, business processes, etc., and provide them to AI business users” (see Section 1.3 above). Therefore, when an AI system introduced by K is managed and provided to internal business divisions or other departments within the company, K may also be regarded as an AI provider.

In the above case, it is first considered that bag seller L bears product liability (under the main text of Article 3 of the Product Liability Act) for the fact that the bag sold contained a foreign object and consumer V was injured. [80, 81]

Furthermore, given that the foreign object originated from a lot within the scope of the inspection contracted to K as inspection service provider, the legal issues are: (i) whether K bears liability to L based on breach of the duty of care of a prudent manager under the business consignment agreement (Article 415 of the Civil Code); and (ii) whether K bears general tort liability to V (Article 709 of the Civil Code).[82]

Additionally, the legal issue is whether AI service developer J bears any liability for the AI's failure to detect the foreign object.

4.1.1 The Position of AI in Hypothetical Case 5

In settings such as the bag manufacturing scenario in Hypothetical Case 5, the conventional approach has been for human inspectors to review X-ray images to check for the presence of foreign objects (see also Reference Judicial Precedent 12 below). In recent years, however, the adoption of high-precision visual inspection AI capable of detecting foreign objects with high probability in the visual inspection of manufactured goods has been advancing.

Such AI possesses the function of effectively processing inspection-related judgments that would otherwise be made by humans in the relevant task, and may qualify as either assistive/supportive AI or dependable/substitutive AI. Where the AI satisfies certain requirements as described below, and where the AI is provided on the premise that it substitutes for part of human judgment and action between the parties, it is considered that this is a category in which the liability of each party should be examined as that of dependable/substitutive AI.

In Hypothetical Case 5, AI is used for primary screening, and by directing images with high uncertainty under AI judgment to human secondary review, the service is able to process inspections with greater precision than conventional manual-only verification in approximately half the time. Settings in which such rapid and large-volume processing becomes possible are considered to be cases in which the necessity of delegating judgment to AI is typically high.

Furthermore, the inspection work in Hypothetical Case 5 is a task that has conventionally been performed primarily by humans. As described in 2.2.1 above, the applicable standard for assessing whether an AI system possesses a sufficient level of precision to substitute for such work

---

[80] For the criteria for assessing defects under product liability, see 4.2.2 below. In cases of foreign-object contamination during the manufacturing process such as in Hypothetical Case 5, it is considered that this constitutes a manufacturing defect.

[81] It is also conceivable that L does not outsource the inspection work to K but instead conducts AI-assisted inspection in-house. In such a case, it is considered that product liability of the manufacturer would arise as a manufacturing defect, without requiring a determination of negligence.

[82] Regardless of the type of claim, since the normative evaluation of what obligation K bore will be the same, the structure of the determination of liability is similar, and no distinction is drawn in 4.1.2 below.

is the work standard of a reasonable person engaged in the same type of task, against which the precision of the AI system in question is to be compared.

Reference Judicial Precedent 12, set out below, provides an indication of the standard of duty of care required for inspection work of the same type as in Hypothetical Case 5. An AI system incorporating an AI model with a very high foreign-object detection rate, such as in Hypothetical Case 5, would be evaluated as meeting a standard that exceeds the level of precision at which a person with a reasonable person's level of attentiveness would typically detect most foreign objects detectable in the same type of work, and thus the rationality of relying on the AI's judgment is more readily recognized. Furthermore, as described in 2.2.1 above, where an AI system achieves a precision level equivalent to or greater than that of a reasonable person only in part of the business process, or where its precision falls below the work standard of a reasonable person under certain conditions, it is also conceivable to secure precision across the entire business process by, for example, incorporating human involvement centered on areas where the AI's detection precision is lower.

Reference Judicial Precedent 12: Standard of Duty of the Contractor under an Inspection Consignment Agreement(Tokyo District Court, March 14, 2014 (unreported))

- **A case in which the plaintiff, having requested the defendant—who owned X-ray inspection equipment—to inspect bags and other products to be delivered to business partners for the presence of metallic foreign objects, sought compensation for damages and return of unjust enrichment on the grounds that the defendant's inspection work constituted a breach of obligation.**
- **Given that the defendant's inspection work under the agreement in this case—which was concluded orally and without any express agreement on the precision of the inspection work—was performed by having a designated employee carefully monitor the screen to determine and confirm the presence or absence of foreign objects,** the content of the obligation cannot be interpreted as requiring, as the obligatory outcome itself, the complete identification and exclusion of all foreign objects. **Rather, it should be interpreted that the agreement was reached on the basis that** the obligation would possess the precision necessary to achieve the purpose of ensuring safety, and that the content of the obligation under the agreement with respect to inspection precision is to possess the level of precision at which a person with average attentiveness would typically detect most foreign objects detectable in the same type of work.

In light of the foregoing, it is considered highly likely that the Visual Inspection AI described above qualifies as dependable/substitutive AI.

4.1.2 Liability of K as AI Provider[83] and AI Business User

Given that the visual inspection AI described above is highly likely to qualify as dependable/substitutive AI, the duty of care of K as AI business user is evaluated, as described in 2.2.2 above,

[83] As discussed in Note 79 above, K also falls under the category of an AI provider under the AI Guidelines for Business.

from the perspective of: (A) whether K appropriately "constructed" the business process incorporating the AI system; and (B) whether K engaged in "operation" while reducing risks to the extent possible.

The specific measures to be taken in relation to the construction and operation of business processes incorporating an AI system will differ depending on the characteristics of each individual AI system, but as described in 2.2.2 above, it is considered that measures within the reasonable range of what is feasible should be taken by reference to the AI Guidelines for Business. K qualifies as both an AI provider and an AI business user, and in relation to the AI Guidelines for Business, it may be required to take measures such as those exemplified in 〈Reference〉 "Measures Envisaged under the AI Guidelines for Business with Respect to Misdetection" on p. 66 below. In relation to Hypothetical Case 5, from the perspective of (A) business process construction, important factors include the appropriate determination of the scope of human involvement and the deployment of personnel experienced in review. From the perspective of (B) operation, important factors are considered to include verifying whether the AI model has degraded and using foreign objects and misdetections encountered in actual operations for continued learning in order to improve the AI model's precision.

Not all of the measures set out in the table on p. 67 below constitute a duty to avoid the consequences; the measures considered to be required will differ depending on individual circumstances such as the content of K's services, the performance and design of the AI, and the allocation of roles with the AI developer.[84] However, it is considered that by constructing governance of the kind described in the table as necessary, this may be evaluated as a circumstance demonstrating that K exercised reasonably feasible care from the perspective of business process construction and operation.[85]

Where K's negligence is not recognized from the perspectives of (A) and (B) above, it is considered that K does not bear a duty to avoid the consequences to the extent of verifying and correcting all AI outputs, even where the AI has reached a misdetection in processing an individual case.

4.1.3 Liability of J as AI Developer

The settings in which J's liability—as the developer and deliverer of the AI—becomes an issue are, as with 3.1.3 and 3.2.3 above, envisaged to include: (1) a setting in which K, having

[84] As discussed in Section 2.2.2(3) above, that the duties of care in design and in explanation required of AI developers and AI providers may differ according to their respective roles within the value chain.

[85] At the study group, it was pointed out that, in Hypothetical Case 5, because inspection contractor K bears responsibility for inspection only within the scope of the work consigned by manufacturer L, the standard of conduct required of K is not considered to be particularly high; accordingly, it should be noted that care is required on this point if the discussion is to be generalized.

compensated V for the damages, seeks indemnification from or damages against J;[86] and (2) a setting in which V or L brings a direct damages claim against J (together with a claim against K).[87]

In scenario (2), which is grounded in tort law, the key question is whether J, in developing and delivering the AI on the premise that it would substitute for part of the judgment in the inspection work, designed the AI to reduce the possibility of misdetection to the extent possible in light of the business objective of ensuring the safety of products subject to inspection, and whether J provided explanations regarding important matters for risk control, such as the scope of areas where human involvement should be retained in the process of AI utilization. With respect to this duty of care as well, whether AI governance was appropriately constructed and implemented may affect the evaluation, and accordingly J is able to reduce the possibility of a breach of duty of care by constructing and implementing, as necessary, the governance exemplified in 〈Reference〉 "Measures Envisaged under the AI Guidelines for Business Regarding Misdetection" on p, 67 below.

Since Hypothetical Case 5 is a case in which the AI ultimately committed a misdetection, it is considered that the design-related measures concerning the AI's detection precision are most likely to be the focus of inquiry with respect to J's liability. By its nature, AI cannot always guarantee intended behavior, and there are developmental limits to detection precision; moreover, the AI was developed and designed within the scope necessary for K's business purpose. Taking this into account, it is considered that, while the work standard of a reasonable person engaged in the same type of task serves as the benchmark as described in (1) above, precision should be evaluated also in light of the content of the agreement between J and K, and if the AI conforms to these standards, it is considered unlikely that J's liability to V as a third party would arise.

〈Reference〉 Measures Envisaged under the AI Guidelines for Business with Respect to Misdetection

The table below organizes, by party in Hypothetical Case 5, the measures envisaged under the AI Guidelines for Business from the perspective of reducing the risk of misdetection of foreign objects.

The specific measures considered to be required will differ depending on the characteristics of the AI and individual circumstances; however, as described in 4.1.2 above and following, taking the measures prescribed by the AI Guidelines for Business as necessary leads to a reduction in the possibility of being evaluated as having acted with negligence. It should be noted that, as described in 2.1.3 above, given the nature of the AI Guidelines for Business as guidelines intended to encourage voluntary initiatives by individual business operators, failure to take the

[86] As with 3.1.3(1) above, although it depends on the specific contractual terms, important factors include, among others: what agreements were made on the performance and quality of the AI; what liability company A will have under the contract if the AI fails to provide the performance, quality, etc. intended under the contract; and what disclaimers were contained in the contract.

[87] In particular, where, as a result of the analysis in 4.1.2 above, K does not bear liability for damages, whether J's liability can be established becomes an important legal issue for V and L.

measures described below does not directly constitute a breach of the duty of care on the part of each party.

			Measures	Guidelines[88]
Inspection Contractor K	AI Provider / AI Business User	Appropriate Construction of Business Processes	● Appropriate use with safety considerations ・Management of models and thresholds used ・Performance verification and continued learning under actual operating conditions ・Verification of changes in model behavior and presence/absence of degradation ・Assessment of the suitability of AI inspection targets and the need for concurrent use of other inspection methods, etc. ● Human Involvement ・Appropriate determination of the scope of review by humans ・Construction of a review system staffed by personnel experienced in inspection, etc.	U-2) i. Appropriate use with safety P-2) i. Risk measures with consideration for human life, body, property, mental wellbeing, and environment, etc. P-2) i. Risk measures with consideration for human life, body, property, mental wellbeing, and environment P-2) ii. Provision conducive to appropriate use, etc.
		Operation to Reduce Risks	● Feedback system upon misdetection ・Feedback to the AI developer regarding significant misdetections ・Cause analysis and corrective measures, etc. ● Recording of usage and transparency ・Preservation of inspection images, judgment logs, and review results ・Ensuring accountability in the event of an incident, etc.	U-6) i. Provision of information to relevant stakeholders U-7) i. Explanation to relevant stakeholders, etc. U-7) i. Explanation to relevant stakeholders, etc. P-6) ii. Provision of information to relevant stakeholders, etc.
AI Service Developer J	AI Developer	Design-Related Measures	● Improvement of precision within developmentally feasible bounds ● Verification of precision and stability in an environment close to actual operating conditions ● Identification of conditions that are difficult to detect (size / position / material, etc.) ● Appropriate design of the scope where human judgment should be applied ● Correction of defects based on feedback from AI business users, etc.	U-2) i. Appropriate use with safety P-2) i. Risk measures with consideration for human life, body, property, mental wellbeing, and environment D-6) i. Ensuring verifiability, etc.
		Explanation-Related Measures	● Clear explanation of performance limitations and residual risks ● Explicit specification of settings and conditions in which human judgment should be applied, etc.	P-2) i. Risk measures with consideration for human life, body, property, mental wellbeing, and environment P-2) ii. Provision conducive to appropriate use, etc.

[88] See the AI Guidelines for Business, pp. 29-41.

4.2 Hypothetical case 6: Autonomous Mobile Robots (AMRs)

AMR Manufacturer M[89] manufactures and sells cart-type Autonomous Mobile Robots (AMRs) that operate in warehouses, factories, and similar worksites. The AMR in question weighs 300 kg, has a payload capacity of 500 kg, and operates at a maximum speed of 1.2 m/s (approximately walking speed), and is capable of rapidly transporting cargo within warehouses and factories in place of conventional carts and forklifts.

Logistics operator N conducted a risk assessment taking into account the performance of M's AMR and the environment of its own warehouse, and decided to purchase and introduce multiple AMRs into a worksite where they would collaborate with human workers. With the introduction of the AMRs, while the AMRs assumed responsibility for transporting large cargo such as pallets (loading platforms), human workers were able to concentrate on detailed tasks, resulting in a significant improvement in overall work efficiency. In addition, during late-night hours when it was difficult to secure human labor, it became possible to continue operating the warehouse at a certain level of efficiency under the supervision of a minimal number of personnel.

The AMR operated safely over an extended period; however, at one point, the following accidents occurred while the AMR was engaged in the transportation of a large volume of cargo.

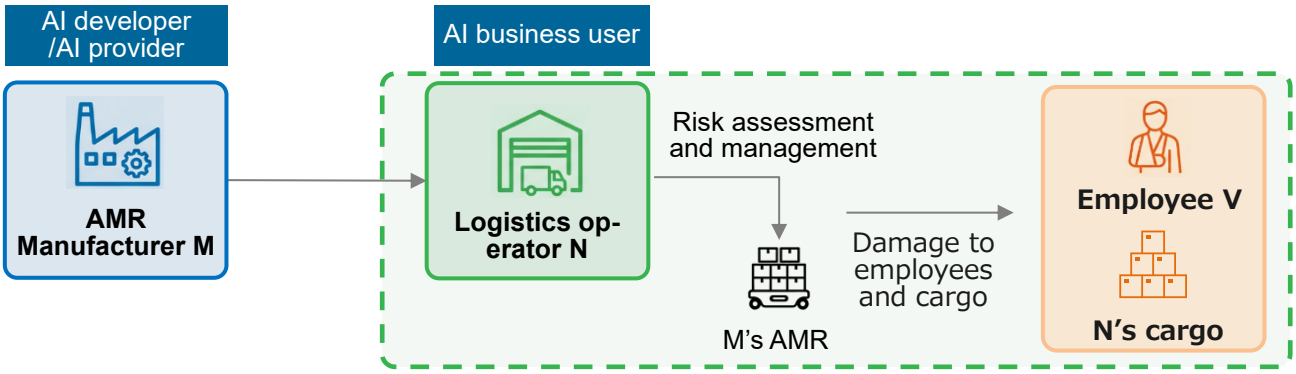
[Case (a)] The AMR collided with employee V, who was in front of the AMR, and injured V. Upon investigating the circumstances of the accident, the facts described in Cases a-1 through a-3 (see 4.2.2 below) were ascertained.

[Case (b)] No bugs existed in the software at the time M delivered the AMR; however, a software update during the course of operations introduced a bug that caused the stopping distance to be calculated as shorter than it actually was under certain conditions, and, as in Case a-1 below, a collision accident occurred that could have been avoided had the proper stopping distance been applied.

[Case (c)] The AMR unit was equipped with a self-diagnostic AI utilizing an LLM, which had the function of performing troubleshooting based on logs and alerts collected from the unit and communicating diagnostic results and remedial measures to on-site workers through a touchscreen panel. At one point, the battery reached a high temperature, and the aforementioned AI detected that the AMR unit had issued a charge-prohibition alarm; however, in its root-cause analysis, it determined that the cause was a "temporary anomaly of the

[89] Because M sells to N AMRs into which M has incorporated its own proprietary control programs, image recognition AI, and other software, M holds both the "status of an AI developer" and the "status of an AI provider" (see the AI Guidelines for Business, main text, p. 5; Annex, p. 13).

temperature sensor" and presented a remedial measure of disabling the charge-prohibition safeguard. When a worker followed the diagnosis, the excessive temperature rise of the battery progressed, and immediately after the battery caught fire and the fire spread to N's cargo stored in the warehouse.



Amid predictions that various machines and robots operating through software including AI will emerge, there is a lack of accumulated judicial precedents regarding the allocation of liability when accidents arise from the behavior of such machines, and predictability remains limited.[90] As one example, the Study Group examined the autonomous mobile robot (AMR), the adoption of which is progressing in factories and similar settings. Such robots differ in characteristics from autonomous vehicles and the like in that they are expected not only to substitute for human labor but also to perform tasks that would have been difficult with human labor alone, such as operation during late-night hours—a benefit beyond mere substitution—and in that the users are business operators who are subject to certain duties of care. These points affect the determinations of liability each party as discussed in 4.2.2 below.

In situations where robots are utilized in business operations as in Hypothetical Case 6, there generally exist the manufacturer that produced the robot (hereinafter collectively referred to as the "Manufacturer") and the user that operates the robot for the purposes of its own business (hereinafter collectively referred to as the "User"), and the question arises as to what liability each party bears.

First, where physical injury to on-site workers has occurred as in Cases (a) and (b), the liability of N as employer and User of the robot (breach of the duty to ensure employee safety or general

[90] It should be noted that the discussions in Hypothetical Case 6 are not applicable to all robots and that attention should be paid to the scope of their applicability; see 4.2.5 below.

tort liability) and the liability of M as Manufacturer (primarily product liability) become issues.[91, 92] Among these, with respect to the User's liability, the characteristics and safety features of the AMR, as well as what design measures and warnings/instructions the Manufacturer implemented, constitute prerequisite factors, and the measures the User should have taken to prevent the accident differ depending on these factors. Accordingly, in 4.2.2 and the sections that follow below, the Manufacturer's liability is examined first, followed by a discussion of the User's liability.

Furthermore, where damage to the User's property has occurred as in Case (c), the liability of Manufacturer M (primarily product liability) becomes an issue.

4.2.1 Safety-Related Technology and Risk Assessment for AMRs

As a prerequisite for discussing the liability of each party involved in the manufacture and operation of AMRs, the technology related to AMR safety and the required risk assessments are organized below.

The functions necessary for safety vary depending on the individual product and the worksite; however, Hypothetical Case 6 assumes an AMR that is equipped with the functions described in (1) and (2) below and has been introduced on the basis of the risk assessment described in (3).

(1) Safety Systems: Functions for Preventing Collisions with Persons and Objects

In terms of standards and specifications, ISO 3691-4 and the corresponding domestic standard JIS D 6802, among others, prescribe the following requirements and PLr (required performance level):

- ✓ Detection sensors such as laser scanners
- ✓ Speed control and monitoring functions
- ✓ Setting of protective fields (stopping zones) based on speed, direction, stopping distance, and other parameters
- ✓ Other safety functions (braking systems, emergency stop devices, alarms, etc.)
- ✓ Person-detection testing under conservative conditions

[91] Conventionally, in the setting of accidents caused by forklifts, for example, the approach taken has been to find the driver's general tort liability and to pursue the employer's employer liability (see Tokyo District Court Judgment, February 16, 2016 (unreported); Osaka District Court Judgment, July 12, 2013 (unreported); Tokyo District Court Judgment, April 7, 2006, *Rodo Hanrei [Laboe Case Law],* No. 918, p. 42; etc.). In contrast, in the setting of operating autonomous mobile robots, because it is not possible to hold a driver liable under general tort for driving decisions made autonomously by the software, the question arises as to the scope of liability borne by each of M and N.

[92] Certain damages such as medical expenses for injured employees are compensated through Workers' Accident Compensation Insurance (labor accident insurance).

(2) Operational Systems: Functions for Smooth Navigation

Functions for achieving smooth navigation include the following, and these functions also affect the behavior and risks of the AMR:

- ✓ SLAM (Simultaneous Localization and Mapping): Creates a map of walls, obstacles, and other features from environmental information acquired by sensors, while simultaneously estimating the position and orientation of the AMR itself.
- ✓ 3D cameras and AI image recognition: Enables the distinction between persons and objects, allowing more flexible operational decisions such as decelerating in advance upon detecting a person.

(3) Risk Assessment by the User

Paragraph (1) of Article 28-2 of the Industrial Safety and Health Act provides that "[a]n employer shall, as prescribed by Order of the Ministry of Health, Labour and Welfare, investigate hazards or harmful factors arising from buildings, equipment [...] or from work activities and other operations [...], and, based on the results thereof, shall endeavor to take necessary measures to prevent dangers to or health impairments of workers, in addition to taking measures under the provisions of this Act or orders based thereon." Accordingly, the User operating the AMR is in a position where it should conduct a risk assessment of the system, including the AMR's onboard mechanisms, peripheral equipment, and operational methods.[93] Specifically, the risk assessment is generally conducted through the following procedures, with reference to JIS B 9700 (ISO 12100), JIS B 9705-1 (ISO 13849-1), and other standards. Where an OHSMS (Occupational Health and Safety Management System) approach has been adopted at the workplace, this risk assessment is positioned as one of its constituent elements.[94]

- ✓ Determination of machinery specifications and limitations (purpose of use, speed, routes, loading, etc.)
- ✓ Identification of hazard sources
- ✓ Risk estimation
- ✓ Risk evaluation
- ✓ Risk reduction[95]
- ✓ Verification, validation, and communication of residual risks

---

[93] See also Japan Industrial Safety and Health Association, *Report on the Project for Promoting Safety Measures for Machinery and Equipment in Response to Technological Innovation*. October 31, 2023. https://www.mhlw.go.jp/content/000830713.pdf (last accessed November 25, 2025), pp. 11 et seq.

[94] It should be noted that the OHSMS approach does not necessarily presuppose certification under ISO 45001 or similar standards, and is considered important regardless of whether certification has been obtained.

[95] Risk reduction methods include inherently safe design measures (such as reducing travel speed), engineering measures (such as adding supplementary sensors or emergency stop buttons), administrative measures (such as establishing navigation rules and educating workers), use of personal protective equipment, and others. See Linx Co., Ltd., *Preventing Failures in AMR Introduction! Part 3: Thinking About 'Safety Measures'—From Legal Interpretation to On-Site Operations* August 20, 2025. https://linx.jp/product/irayple_amr/article/009/ (last accessed November 25, 2025).

✓ Recording of implementation status, change management, and periodic review

4.2.2 Case (a)

(1) The Position of the Autonomous Navigation in Case (a)

As described in 4.2.1 above, an AMR is constituted by the integration of diverse software and hardware, and AI using machine learning is one part thereof. Accordingly, this section examines whether the autonomous navigation as a whole satisfies the same requirements as dependable/substitutive AI and qualifies as a system to which operational decisions may be entrusted (hereinafter in this section referred to as a "dependable/substitutive system").

The AMR is equipped with the autonomous navigation system, which by its nature is not suitable to classification as an assistive/supportive system; moreover, as set forth below, it is considered that it may also qualify as a dependable/substitutive system as a matter of legal evaluation.

The AMR is capable of autonomously navigating within the worksite and thereby smoothly and rapidly transporting cargo within the factory in place of conventional carts and forklifts. The benefits of introducing AMRs depend on the individual usage environment; however, as in Hypothetical Case 6, it is considered that the AMR contributes to a significant improvement in work efficiency through the division of labor between humans and AMRs, as well as to the performance of operations that would have been difficult with human labor alone, such as warehouse operation during late-night hours, and accordingly, the necessity of entrusting human transportation work to AI is recognized.

Furthermore, with respect to the safety of the autonomous navigation, when transporting large volumes of cargo using machinery and the like in worksites such as warehouses, there is a risk to the life and physical integrity of workers—significant legal interests—through collision accidents and the like, and a high degree of care is generally required of those engaged in similar work.[96] In evaluating whether the AMR possesses a level of safety equal to or greater than that required for similar work, since the AMR is fundamentally a robot that does not presuppose human operation, it is not necessarily easy to compare the quality and accuracy of its decisions with human work standards in accordance with the criteria set forth in 2.2.2 above. On the other hand, as one point of comparison, when compared with human-operated forklifts that have been used for transport work in warehouses and similar settings, the AMR has characteristics such as speed limitations for risk reduction and the incorporation of designs such as those described in 4.2.1(1) and (2) above for ensuring and improving safety. In addition, since the User is required under the Industrial Safety and Health Act to conduct a risk assessment (see 4.2.1(3) above), and introduction of the AMR presupposes that risks are reduced to an acceptable level through countermeasures including, as necessary, restrictions on the AMR's speed, usage environment, usage

[96] As one example, see the judicial precedents cited in Note 91 above concerning accidents that occurred during cargo transport by forklift.

conditions, and the like, it is considered that, with respect to navigation within the scope of usage environment restrictions based on the risk assessment, the AMR is capable of demonstrating safety equal to or greater than that of human-operated forklifts.[97]

Based on the foregoing, the AMR may qualify as a dependable/substitutive system. In its usage context, the Manufacturer is required, under the Product Liability Act, to take the necessary design and warning/instruction measures to enable the AMR to demonstrate the aforementioned level of safety, on the premise that decisions regarding navigation in the course of transport operations are entrusted to the AMR's autonomous behavior, and the User bears a duty of care with respect to the proper establishment and operation of work processes incorporating the AMR.

In this regard, since the AMR makes flexible navigation decisions and actions in response to on-site environmental conditions, it is not easy to determine the extent of the safety that is intended, and as discussed in (2) below, it is considered that a comprehensive evaluation incorporating multiple perspectives is necessary.

(2) Framework for Determining Manufacturer's Product Liability in AMR Collision Accidents

(A) Approach to Products That Operate Autonomously Through Software

As described in 2.1.2 above, in a claim for damages under the Product Liability Act, the requirement is that the product manufactured, processed, imported, or otherwise handled by the manufacturer, etc. has a "defect" (Article 3 of the Product Liability Act). Generally, defects are classified and explained in the following three categories:

Manufacturing defect(s)[98]	Cases where the product was not manufactured in accordance with its design or specifications and lacks safety
Design defect(s)	Cases where, because sufficient consideration was not given to safety at the design stage of the product, the product lacks safety as a result

[97] For specific methods of conducting safety evaluations in comparison with human-operated machinery, see (2) below. Where it is not straightforwardly possible to compare against the work performance standard of a reasonable person, it is also conceivable, as noted in Note 32 above, to take into account as supplementary considerations: (A) a balancing of the benefits of operational efficiency improvement and the like against the risks (the likelihood of rights infringement or damage occurring and its gravity); and (B) compliance with requirements under regulatory law. From the perspective of (A), since AMR operation involves significant legally protected interests such as workers' physical wellbeing, the application of a balancing standard should be interpreted conservatively; however, taking into account the results of the User's risk assessment and similar factors, if the risks that may increase or the damage (should it occur) as a result of introducing the AMR are minimal, it is considered that it may be permissible to give supplementary consideration to the benefits of introduction. From the perspective of (B), it is considered that conducting a risk assessment under the Industrial Safety and Health Act and taking measures to reduce risk to an acceptable level as necessary may be taken into account as one element in support of a finding that the required safety has been attained.

[98] It has been pointed out that, since software installed in robots and the like is copied in its entirety, manufacturing defects are difficult to conceive of (Technova Inc. (2020) *Research Report on Civil Liability and Social Acceptance of Automated Driving*. p. 12).

Warning/instruction defect(s)	Cases where, with respect to a product that has risks that cannot be eliminated in light of its utility or efficacy, the manufacturer fails to provide appropriate information for the consumer to prevent or avoid accidents arising from the manifestation of such risks

With respect to robots such as AMRs, in addition to issues concerning defects and design flaws in the unit itself (e.g., risk of fire) that arise in the same manner as with conventional products, the question of how determinations of defect should be made with respect to accidents caused by software behavior becomes an issue. In particular, with regard to design defects, since products such as robots that operate autonomously through software including AI are by their nature expected to exhibit flexible behavior, the question of what behavior would be evaluated as deviating from the required design, and consequently as lacking "the safety that the product should normally have," is difficult.

With respect to autonomous vehicles as one example of autonomously operating machines, primarily the following approaches exist, which may also serve as references for determinations of defect concerning other machines and robots: [99], [100], [101]

	Overview and Basis	Challenges in Applying to Products Other Than Autonomous Vehicles
(i) Approach Based on Overall/Statistical Safety	● An approach that uses as its criterion whether the autonomous decisions of the product such as an autonomous vehicle are probabilistically and statistically safer	● There may be situations where it is difficult to compare statistical safety, and for products such as AMRs that are still in the process of widespread adoption, it is not uncommon for data collection to be difficult

[99] Under both approaches, it is considered that, for the time being, the safety of human-operated machinery will serve as the comparison benchmark; however, it is also anticipated that, as technology advances, more safety-capable systems of the same type may become the benchmark (see Note 31 above).

[100] When examining safety under each approach, compliance with safety standards under regulatory laws, technical standards, and specifications also constitutes a strong consideration. Regarding the relationship between compliance with regulatory law and determinations of defect, see Consumer Affairs Agency (September, 2018) *Article-by-Article Commentary on the Product Liability (PL) Act.* https://www.caa.go.jp/policies/policy/consumer_safety/other/product_liability_act_annotations (last accessed November 25, 2025), p. 20. With respect to compliance with technical specifications, there is a judicial precedent holding that a child seat that complies with Japanese Industrial Standards may be presumed, absent special circumstances, to be a product free from defects in terms of its restraining function (Hiroshima District Court (Miyoshi Branch) Judgment, February 19, 2007 (unreported)), and it is considered that standards establishing detailed safety criteria may constitute one important consideration in the determination of defects; on the other hand, there is also a judicial precedent holding that failure to conform with JIS standards does not constitute a defect, since JIS standards are intended to standardize the quality of industrial products and the like (Tokyo District Court Judgment, November 27, 2007 (unreported)), and ultimately, individual case-by-case examination is required.

[101] It should be noted that where the existence of a defect could not have been recognized in light of the state of scientific or technical knowledge at the time of delivery of the product, no liability for damages arises (item (i) of Article 4 of the Product Liability Act—the development risk defense. "Knowledge" refers to all knowledge that has been established to a degree capable of affecting the determination of defect, and refers not to the knowledge of a specific person but to the aggregate of knowledge that objectively exists in society (Consumer Affairs Agency, cited in Note 100 above, p. 38)). While there are no judicial precedents to date in which the development risk defense has been recognized, it is conceivable that, in the future, cases may arise in which a finding of development risk is made with respect to AI—a scientific technology that inherently involves uncertainty—on the grounds that the defect could not have been recognized in light of the state of scientific or technical knowledge at the time of delivery.

	compared with human decisions and actions[102] ● The probability of accident occurrence is considered to be one of the principal factors in determination of defect[103]	
(ii) Approach Based on Safety in Individual Circumstances	● An approach that evaluates whether the product demonstrated the level of safety it should have exhibited, taking into account the individual circumstances in which the accident occurred ● In the context of autonomous vehicles, approaches have been advocated that demand safer driving behavior than that of an average driver[104] ● From the perspective of protecting consumer expectations, a value judgment may be made that safety in individual circumstances should be evaluated	● In the case of machines such as AMRs that do not presuppose human driving operation, it may become difficult to evaluate safety in individual circumstances through comparison with human operation

These criteria may also be applied to AMRs; however, the following difficulties also arise.

First, in the approach focusing on the overall/statistical safety of AMRs (item (i) in the table above), a possible method would be to compare, for example, the statistical information on industrial accidents across conventional similar worksites as a whole with the accident rate per operating hour of AMRs manufactured by M.[105] Additionally, as an eclectic approach that maintains the overall/statistical framework while focusing more on safety in individual circumstances, a possible method would be to compare the historical industrial accident rate in similar work at the worksite where the AMR was introduced (N's warehouse in Hypothetical Case 6) with the accident rate per operating hour after the introduction of AMRs manufactured by M. In either approach, the premise is the collection of a statistically significant volume of data for comparing and evaluating safety; however, since AMRs are still in the process of widespread adoption, it is not easy to collect sufficient data at the time of preparation of this document. Whether "the safety that the product should normally have" can be evaluated solely on the basis of the above criteria is considered to be a matter for future consideration. It should be noted that even in cases where it is difficult to evaluate "the safety that the product should normally have" solely through statistical evaluation, since the probability of accident occurrence is an important factor, the fact that the

---

[102] Toshima Koji et al. (eds.) (2024). *Jidounten·MaaS Business no Homu (Dai 2-han) [Legal Practice in Autonomous Driving and MaaS Business (2nd ed.)] (in Japanese)*. pp. 70–71. Chuo Keizai Sha

[103] Consumer Affairs Agency, cited in Note 100 above, p. 9.

[104] Urakawa, Michitaro (2017). Civil Liability and Automated Driving (in Japanese). *NBL, No. 1099,* p. 34. As one specific example of the determination method, a method has been presented of using EDR (black-box) records to compare the reaction times of an "average natural person driver" and an autonomous vehicle in terms of their ability to foresee and avoid a dangerous condition at the time of the accident in question; since the reaction time of an "average natural person driver" is also expected to be ascertained through statistical methods, it is considered that the statistical perspective of (i) in the table above may also be used in connection with the approach of (ii) in the table.

[105] This accident rate is closely related to insurance premiums (Mark A. Geistfeld (trans. Segawa, Nobuhisa et al.) (2025). Artificial Intelligence Systematizes U.S. Tort Law (in Japanese). *Waseda Law Review, Vol. 100, No. 4,* pp. 206–207).

product is statistically safe is considered to be one circumstance that may be taken into account in negating a defect.

Furthermore, with respect to the approach based on safety in individual circumstances (item (ii) above), since the AMR fundamentally does not presuppose human operation, it is conceivable to compare it with forklifts that have been used for similar work and to compare the safety with the operations of an ordinary driver that would be assumed in the scenario of the actual accident. Specifically, it is conceivable to evaluate whether it is highly probable that the accident could have been avoided through the decisions and operation of a human driver, or whether it is highly probable that the accident could not have been avoided in any event. However, since forklifts and AMRs are machines with different structures and the external environment, including the scope and content of safety management measures that users should implement. Therefore, comparisons with the operation of forklifts should be regarded merely as one rough guide in the determination of "the safety that the product should normally have," and there may also be cases where comparison is difficult.

The above framework constitutes an important perspective in relation to determinations of defect for AMRs; however, it is considered difficult to evaluate whether "the safety that the product should normally have" has been satisfied on the basis of any single criterion alone. Accordingly, it is necessary to evaluate safety comprehensively while also applying the traditional framework for determinations of product liability that has been applied to conventional industrial machines and the like.

(B) Approach Regarding Conventional Industrial Products

With respect to the determination of defects for conventional industrial products and the like, the following factors have been considered: (i) characteristics of the product, (ii) the ordinarily foreseeable manner of use, and (iii) the time of delivery and other circumstances (paragraph (2) of Article 2 of the Product Liability Act). The analysis on these factors[106] with respect to AMRs is as follows:

Indicator	Perspective
Labeling of the product (relating to item (i) above)	● While the risk of collision and the like is relatively easy to perceive externally, the technical design content that affects safety and the situations in which unexpected behavior is likely to occur are difficult to ascertain externally, and clear explanations of risks are more readily required
Utility/usefulness of the product (relating to item (i) above)	● The AMR is capable of transporting large volumes of cargo by operating with agility and speed, compared with work using carts and the like ● If sensors and the like are designed with excessive redundancy such that stopping occurs too frequently, the fundamental purpose of smooth operation itself is impeded

[106] Consumer Affairs Agency, cited in Note 100 above, p. 7. While other considerations also exist, those considered particularly important in relation to AMRs have been extracted here.

Availability of alternative design / cost-effectiveness (relating to item (ii) above)	● While it is possible to improve safety through customization such as adopting multiple layers of redundant design by combining different types or multiple sensors, the safety design required varies depending on the usage environment, and specifications are determined on the basis of the User's risk assessment, with costs determined on the basis of those specifications
Possibility of damage prevention by users (relating to item (ii) above)	● The usage environment varies depending on the conditions of each individual factory or warehouse and is under the User's control; accordingly, the robot's speed control and other parameters are customized on the basis of the User's risk assessment (see 4.2.1 above), and layout adjustments and changes are also made ● Active accident avoidance measures are intended to be taken through the design of the usage environment and education/instruction of workers
Reasonably foreseeable manner of use (relating to item (ii) above)	● Although certain dangerous behavior and operation deviating from the intended usage environment are foreseeable, collaboration with persons who have received training on autonomously navigated vehicles is assumed, and reasonable responses by workers are also more readily expected
Technical standards at the time of delivery (relating to item (iii) above)	● At the time of preparation of this document, it is generally the case that technologies such as those described in 4.2.1 above are employed ※Note: For robots such as AMRs, it is common to improve quality and safety through software updates are contemplated by their nature; for determinations of defect taking this point into account, see 4.2.4 below.

Among the items in the above table, a particularly important characteristic is that the smoothness of navigation and the level of safety required of AMRs vary depending on the individual usage environment into which they are introduced, and additional design features and the like are selected on the basis of the risk assessment by the User who managed the human and physical resources of the worksite. This point should be given weight in determinations of defect as described in (iii) below.

(C) Direction of Determination of Defect in Case (a)

Based on items (A) and (B) above, it is considered that the determination of defect should proceed along the following lines:

(a) Design Defect(s)[107]

First, as described in (A) above, data on the AMR's statistical accident rate may serve as a factor in considering whether the necessary level of safety was met; in addition, assuming hypothetically that a human-operated forklift had been used at the time of the accident, whether the

[107] It has been pointed out that, in the determination of defects in autonomous vehicles, the safety that the vehicle should normally have should be evaluated taking into account the fact that a driver's reliance on a system such as automatic emergency braking may result in a lower level of attentiveness than that of a driver of a vehicle not equipped with such a system (Kubota, Mitsumi (2018) Liability of Dealers and Manufacturers in the Context of Autonomous Driving, Fujita, Tomoyuki (ed.) *Jidounten to Ho [Autonomous Driving and Law] (in Japanese).* p. 174. Yuhikaku). However, in the setting of AMR operations, since the User is a business operator and is in a position requiring a more cautious judgment in relation to ensuring workplace safety, it is considered that the influence of the User's reliance on the system on the determination of "the safety that it should normally have" is limited.

accident could have been avoided through human decisions and actions also serves as one benchmark.

However, since it is considered difficult to conclusively determine the existence of a defect based on these factors alone, based on the factor analysis described in ② above, the criterion should be whether a reasonable alternative design capable of avoiding the accident existed, taking into account the actual usage environment and the specific design of the AMR (weight, speed, stopping distance, etc.) in the case in question. In examining the availability of an alternative design, the following points should be taken into consideration:[108]

- With respect to risks in the ordinary course of use in the intended usage environment, the availability of an alternative design is more broadly inquired into; and even for reasonably foreseeable misuse or deviations from the intended usage environment, a certain degree of fail-safe and other design considerations are required. Where the intended specifications and design failed to function due to a serious bug or other defect, since an alternative design free from such a defect is normally possible, there is a high possibility that a design defect will be found.
- On the other hand, with respect to accidents attributable to the usage environment or usage methods, in light of the fact that the worksite is under the User's control and that a risk assessment is a prerequisite for introduction, a design defect is unlikely to be found, and a failure by the User to take accident avoidance and prevention measures is more readily recognized.
- Furthermore, designs that were agreed upon and determined on the basis of the User's risk assessment should also be respected in the safety evaluation, and alternative designs premised on costs greatly exceeding those agreed upon are unlikely to be recognized as reasonable.

(b) Warning/Instruction Defect(s)

As a prerequisite for the User to discover hazard sources through risk assessment and to take reasonable avoidance actions while appropriately utilizing the safety functions, clear explanations from the Manufacturer to the User regarding fundamental functions, performance limitations, and the like are required.

For example, information such as the detectable range of sensors and stopping distances is considered important for appropriate accident avoidance actions. Additionally, when the AMR is heading directly toward a person or wall and the designed stopping function cannot stop the AMR in time, it may be necessary to stop the AMR externally using the emergency stop button, and the situations in which such an emergency stop is necessary and the methods of operation are also

[108] Detailed safety standards are established in ISO and JIS standards and serve as a provisional benchmark; however, the ultimate evaluation depends on whether a reasonable alternative design capable of avoiding the individual accident in question can be recognized.

considered important for accident avoidance. In providing explanations, in addition to accuracy, ease of understanding by the User is also an important consideration, and there may be cases where it is more important to provide explanations at an operational or functional level that is intuitively understandable—such as "given the AMR's design, at what point will an obstacle cause the AMR to stop?" and "at what angle will obstacles make stopping difficult?"—rather than providing technical details.

(3) M's Product Liability

In Case (a), various patterns of collision are conceivable, and M's product liability is determined differently for each pattern.

[Case a-1] Accident Attributable to a Defect in the Development Process

V was walking from the right side of the AMR and entered its path. M's AMR was equipped with a protective stop function compliant with JIS, which was designed to calculate the stopping distance based on speed and other parameters. However, because there was a bug in the software that caused the stopping distance to be calculated as shorter than it actually was under certain conditions, the stop timing was delayed and a collision occurred. Had the originally intended stopping distance been applied, it would have been possible to avoid the collision.

In Case a-1, from the perspective of an overall/statistical safety assessment, if there are circumstances such as a declining trend in the rate of industrial accidents at the site where the AMR was introduced (N's warehouse) compared to before the introduction of the AMR, this may tend to support a conclusion negating the existence of a defect. On the other hand, as a comparison with a human-operated machine, assuming hypothetically that a human-operated forklift had been used rather than an AMR, there is a possibility that the relatively slow movement of V walking and entering the path ahead would be evaluated that the driver could have visually detected and stopped or taken other measures before the collision. In such a case, this could serve as one factor supporting a finding of defect.

On that basis, in Case a-1, nevertheless the AMR is designed to be capable of avoiding the accident in question, it occurred because the software bug prevented the intended protective stop function from being exercised. This function constitutes a fundamental element of the AMR's basic safety, and it is generally considered difficult for the User to take measures to avoid the accident in anticipation of such function not being exercised. Accordingly,[109] an alternative design free from such a defect is required and there is a high possibility that a design defect will be found.

[109] As software becomes more complex, achieving a state entirely free of bugs becomes unrealistic and it is not the case that all bugs are evaluated as lacking "the safety that the product should normally have"; rather, cases in which an important safety function fails to perform as in Case a-1 are considered to be the typical examples of a design defect.

[Case a-2] Accident Attributable to Usage Environment or Usage Methods

M's AMR was susceptible to reduced sensor sensitivity due to surface contamination; accordingly, M had explained to N the risk along with the necessity of periodic maintenance and the maintenance methods. However, N neglected maintenance during the course of use and contamination had accumulated on the unit. As a result, the detection, alarm, and braking for a worker who was working on the AMR's path without noticing the AMR were delayed, and a collision occurred.

In Case a-2, from the perspective of an overall/statistical safety assessment, it is also conceivable to compare the accident rates before and after the introduction of the AMR, as in the above case. Additionally, as a comparison with a human-operated machine, Case a-2 involves a situation where the AMR's sensor sensitivity had decreased because User N had neglected maintenance. In order to make a meaningful comparison with human-operated machines in such a case, it would be necessary to set identical conditions; for example, one could evaluate whether the above accident could have been avoided if a forklift with a contaminated windshield and impaired visibility had been used. However, verifying the possibility of avoiding an accident under such a hypothetical scenario is not necessarily easy, and Case a-2 is considered to be a case where comparison with a human-operated machine is fundamentally difficult.

From the perspective of the appropriateness of the design, while measures to prevent all accidents caused by M's usage contrary to the N's explanation are not required, the above is considered to constitute a reasonably foreseeable misuse, and the question arises as to whether a certain fail-safe mechanism, such as a system for detecting and warning of abnormalities in sensor sensitivity, was provided. Where such a mechanism is provided, the possibility of constituting a defect on the grounds that a further alternative design should have been adopted is low.[110]

Furthermore, where M has explained the risk of reduced sensor sensitivity due to surface contamination of the unit and the necessary maintenance methods, N was in a position to reasonably foresee the occurrence of the above accident and was also capable of implementing countermeasures such as maintenance. Where such information for accident avoidance has been provided to the User, a warning/instruction defect will also not be found. Additionally, it is considered that the fact that such warnings/instructions were given enhances the possibility of accident avoidance by the User and operates as a circumstance negating a design defect.

[110] It should be noted that where multiple sensors or cameras are used in combination, it is conceivable that the accident might have been avoidable; however, as described in 4.2.3(2)(A) above, the usage environment and design content agreed upon at the outset are respected, and a design premised on costs that significantly deviate therefrom is difficult to recognize as a reasonable alternative design.

[Case a-3] Accident Arising from the Interplay of Design Limitations and Usage Methods

In addition to the JIS-compliant protective stop function, M's AMR was equipped with a 3D camera and image recognition AI as a function for improving navigation efficiency, detecting workers approaching from the side and issuing early warnings while decelerating. However, the AI had detection limitations and M had explained to N the risks including the fact that detection accuracy would decrease for unusual silhouettes.

V was moving while carrying a large stepladder on their shoulder in a sideways position to respond to a temporary trouble at the worksite, preventing the AI installed in the AMR from detecting V's silhouette as a person, issuing a warning and decelerating. Subsequently, when V suddenly entered the AMR's stopping distance, the protective stop function was activated, but the stop was not in time and a collision occurred.

In Case a-3, from the perspective of an overall/statistical safety assessment, it is also conceivable to compare the accident rates before and after the introduction of the AMR, as in the above case. Additionally, as a comparison with a human-operated machine, assuming hypothetically that a human-operated forklift had been used rather than an AMR, there is a possibility that it would generally be evaluated as difficult to avoid a collision in advance where V suddenly entered the path ahead, and in such a case, this could serve as one factor supporting an evaluation that no defect existed.

From the perspective of the appropriateness of the design, the important elements in Case a-3 are the protective stop function and the image recognition function. Analysis on these functions is as follows:

■ **Regarding the operation of the protective stop function:**

Where such a safety design failed to function properly due to a software bug or the like, as in Case a-1 above, there is a possibility that a design defect will be found on the grounds that an alternative design should have been adopted. On the other hand, where, as in Case a-3 above, the safety design was functioning but the accident occurred because V suddenly approached, this tends toward negating a defect.

■ **Regarding the operation of the image recognition function:**

The question arises as to whether the developer took performance improvement measures that were reasonably feasible as a developer, such as improving AI accuracy to the extent reasonably possible in development and verifying accuracy and stability in an environment close to actual operations. However, given that it is generally difficult to improve AI accuracy for exceptional situations, and given that the above protective stop function is provided as a

safeguard against AI misdetection, the possibility that a design defect will be found on the grounds that the AI failed to detect an unusual silhouette, as in Case a-3, is low.[111]

On that basis, where it is found that the User was capable of taking accident avoidance measures such as avoiding the transport of cargo that is prone to cause misdetection or suspending AMR operations during responses to temporary troubles, this tends toward negating a design defect.

Furthermore, from the perspective of warnings/instructions, M should explain, within a reasonable scope, the outline of the technical configuration relevant to risks of navigation (e.g., that the AMR is equipped with a protective stop function as well as a 3D camera and image recognition AI). In addition, depending on the User's level of understanding and other factors, explanations regarding the situations in which the AI is prone to misdetection and the risks in the event of misdetection may also be required.

(4) Liability of N and V

(A) Liability of N as User

N, being in a position of owing a duty to ensure the safety of employees under the employment contract with V, bears an obligation to organize appropriate human and physical resources for the prevention of hazards and to provide safety education.[112] Furthermore, upon introducing the AMR, N is required by law to conduct a risk assessment (Article 28-2 of the Industrial Safety and Health Act). Taking these together, it is considered that N bears an obligation to take reasonable measures for accident avoidance, including measures and organizational arrangements determined to be necessary through risk assessments at the time of introduction and during the course of operations, as well as instruction and education of employees, in order to reduce risks.

In relation to each of the above cases, in a case where an accident occurred because the functions intended as the AMR's basic safety performance were not exercised due to a software bug (Case a-1), it would be difficult to adopt safety measures in anticipation of such a defect. It is considered that this would not be evaluated as a breach of the duty to ensure employee safety in relation to the occurrence of the accident.

By contrast, the failure to properly maintain the AMR unit (Case a-2), and the failure to take accident avoidance measures such as avoiding the transport of cargo prone to causing misdetection or suspending AMR operations during responses to temporary troubles (Case a-3), may be evaluated as a breach of the duty to ensure employee safety by User N.

---

[111] It should be noted that, since it is inherently difficult to guarantee that image recognition AI will always achieve the desired level of detection, failure to detect as intended is difficult to evaluate as a software bug.

[112] Supreme Court Judgment, February 25, 1975, Minshu, Vol. 29, No. 2, p. 143; Supreme Court Judgment, May 27, 1983, Minshu, Vol. 37, No. 4, p. 477; see also Okuda Masamichi (ed.) (2011). *Shin Chushaku Minpo [Japanese civil law annotated, new edition] (10) Saiken no Mokuteki· koryoku [Object and Effect of claim] (2) (in Japanese)*. pp. 129–131 [Kitagawa Zentaro = Siomi Yosio]. Yuhikaku, etc.

(B) Negligence of Victim V

Even where M or N bears liability for damages, if negligence on V's part is found in relation to the accident, the amount of damages may be reduced through apportionment of negligence (paragraph (2) of Article 722 of the Civil Code). "Negligence" in the context of apportionment of negligence is not limited to a legal duty of care (a duty to avoid results premised on foreseeability), but is sufficient if it constitutes inattention on the part of the victim that can be evaluated as having contributed to the occurrence of the accident or the expansion of the damage;[113] for example, deviations from manuals or procedures established by N may be subject to apportionment of negligence.[114]

4.2.3 Case (b)

The basic facts of Case (b) are the same as those of Case a-1; however, the difference is that at the time of delivery no defect causing the accident existed, and the defect arose from a software update after delivery of the AMR. In this case, the question arises as to whether only general tort liability (Article 709 of the Civil Code) applies, or whether product liability may also apply.[115]

V may bring a claim against M based on at least general tort liability (Article 709 of the Civil Code). V must allege and prove M's negligence (i.e., what duty to avoid results M bore, premised on foreseeability of the accident), and if successful in such allegation and proof, V will be able to obtain compensation for damages.

On the other hand, while Case (b) expressly identifies the cause of the accident for the purpose of clarifying the issues, there are also cases where the cause of the accident is not apparent. Given that important materials relating to the product are concentrated in the possession of M and N, and that it is conceivable that V may face difficulties in investigating and identifying the cause to allege and prove negligence, V may also consider bringing a claim based on product liability in addition to general tort liability. In a claim based on product liability, V would allege and prove defect rather than negligence, being able to utilize the doctrine of de facto presumption of defect (see 5.1.3 below).

However, since the reference time for determinations of defect under product liability is understood to be the time of delivery of the product,[116] the question arises as to whether a defect

[113] Otsuka Tadashi (ed.) (2022). Shin Chushaku Minpo [Japanese civil law annotated, new edition] (16) Saiken [Claims] (9) (in Japanese), p. 490 [Kashimi Yumiko]. Yuhikaku.

[114] Where V's actions contributing to the accident were so unexpected, it is conceivable that the causal relationship between the AMR's defect or N's breach of the duty to ensure employee safety and the rights infringement may be denied.

[115] Circumstances may arise with technological advancement in which updates become necessary, and if M fails to perform the necessary updates in such a case, general tort liability is considered to become an issue.

[116] Consumer Affairs Agency, cited in Note 100 above, p. 7.

arising from a software update after delivery may be considered in the determination of defect. On this point, two views were presented in the Study Group.

One view holds that, while a defect caused by a software update cannot in principle be evaluated as a defect at the time of delivery, where a latent cause of the defect from the software update (i.e., circumstances establishing the necessity that a software update must be performed) existed at the time of delivery of the product, it may be evaluated that a defect existed at the time of delivery. The rationale is that, given that the reference time for determination of defect is the time of delivery, and that the meaning of "delivery" is generally interpreted in accordance with the concept of delivery of movables under the Civil Code (Articles 178, 182, 183, and 184 of the Civil Code)[117], the scope of defects must be subject to limitations. Under this view, in Case (b), the defect arising from the software update would be considered in the determination of defect only where, at the time of delivery of the AMR, a defect existed that required remediation through a software update.

The other view holds that, with respect to products for which software updates are contemplated at the time of delivery, the effects up to the time of the final update may be considered in the determination of defect. The rationale is that software updates for improving the efficiency and safety of navigation are naturally contemplated at the time of delivery, that such updates are performed by the Manufacturer itself as a management activity, and that the internal structure governing the robot's behavior may be evaluated as having been newly delivered through the software update. Under this view, the defect arising from the update in Case (b) would also be considered in the determination of defect. While the issue of consistency with the concept of delivery of movables under the Civil Code arises, it is considered that, for types of AI in which improvement of quality and safety through software updates is presupposed—unlike the typical products contemplated at the time of enactment of the Product Liability Act—the need for a flexible interpretation of the meaning of "delivery" such as the above is increasing.[118]

4.2.4 Case (c)

These days, examples have emerged in which AI such as language models are installed in robots to perform troubleshooting, as in Case (c). Where an accident results from a worker's actions taken on the basis of the output of such a diagnostic function, User N may seek damages

[117] Consumer Affairs Agency, cited in Note 100 above, pp. 18–19.

[118] It should also be noted that, at the Study Group, as one interpretive approach premised on the fact that the product is intended to be used while software updates continue to be applied, there was a suggestion of treating delivery as occurring continuously and in real time throughout the course of the robot user's use. Under this interpretation, if a determination of defect based on the criterion of statistical safety as set forth in 4.2.2 above reveals that there is a problem with statistical safety at a given point and a software update is not performed, a "defect" would be found without the need to establish negligence. Conversely, where the manufacturer was diligently applying software updates but a problem with statistical safety arising from factors that could not be recognized in light of the state of scientific or technical knowledge at the time is discovered, the development risk defense (see Note 101 above) would be applicable.

from Manufacturer M. The questions arise, first,[119] whether product liability applies, and second, how the determination of defect should be conducted.

(1) Applicability of Product Liability

As described in 2.1.2 above, where software is incorporated into a product, the behavior of such software is also considered in the determination of defect, and Cases (a) and (b) clearly fell within this category.[120] On the other hand, the software and UI configuration of the above troubleshooting function may take multiple forms; for example, the AI that performs inference may be installed on a server, host computer, or other device independent of the AMR, and evaluations may differ as to under what circumstances it can be said that the software is incorporated into the product.

The above self-diagnostic AI is provided by Manufacturer M as part of a system integrated with the AMR, as one element of the functionality for using the AMR safely. The output of the AI is contemplated as a characteristic of the product to be displayed through the unit's touchscreen panel, and the interface indispensable for providing the diagnostic function is installed on the product. Taking into account the system integration and functional connection between the AI and the AMR, if there is any error in the AI's output, it may be evaluated as a defect in a component of the AMR; and even where the inference itself is performed on a device separate from the AMR unit via the Internet, LAN, or other network, it is considered that the application of product liability is not negated.

(2) Framework for Determination of Defect

(A) The Position of the AI in Case (c)

The above self-diagnostic AI may be used on the premise that the AI's accuracy has been sufficiently enhanced such that the User and its employees follow the AI's output (dependable/substitutive AI), enabling smooth warehouse operations even under reduced-staffing conditions such as during late-night hours; alternatively, it may also be contemplated for use solely as an aid or support for the decisions of the User and its employees (assistive/supportive AI). In this regard, given that the User does not necessarily possess expertise in robot maintenance, there is a possibility that the User may become excessively dependent on the AI's output, and it is considered important to clearly explain in the instruction manual or other documentation as to which category the AI is to be used under.

The framework for determination differs depending on which category applies, as follows:

(B) Where the AI Qualifies as Dependable/Substitutive AI

---

[119] If, hypothetically, a manufacturer O separate from M manufactured a touchscreen panel equipped with AI, and M incorporated it into the AMR, then O's product liability with respect to the touchscreen panel, and a defense under paragraph (2) of Article 4 of the Product Liability Act, may become issues.

[120] See Consumer Affairs Agency, cited in Note 100 above, p. 2.

In providing a dependable/substitutive AI, M is required to take design measures to achieve the desired accuracy. With respect to the standard required, while it was difficult in Case (a) to evaluate navigation safety solely through comparison with human operation, since troubleshooting in Case (c) is a domain that has traditionally been handled by human technicians, it is conceivable to evaluate the safety of the AI's output by comparing it with the decisions and actions of an ordinary technician confronting the above situation. In the above case, since the AI recommended dangerous actions such as disabling a safeguard after taking into account an alert concerning the battery temperature rise, there is a high possibility that the AI's output would be construed as lacking safety compared with an ordinary technician.

Furthermore, as N does not necessarily possess expertise in robot maintenance, where access controls and the like to prevent employees from making dangerous configuration changes such as disabling safeguards have not been implemented, this may be evaluated as a design defect.

(C) Where the AI Qualifies as Assistive/Supportive AI

In providing an assistive/supportive AI, M is required to provide warnings/instructions regarding the functions, usage scenarios, usage methods, and the like, on the premise that the output may contain certain errors. So long as such explanations are provided, the possibility of a warning/instruction defect being found is low. Furthermore, since, based on such explanations, N is expected to verify and correct the AI's output even if the AI produces erroneous output, the fact that the AI produces such erroneous output as described above, it would not be considered a design defect.

However, as with (B) above, where access controls and the like to prevent employees from making dangerous configuration changes such as disabling safeguards have not been implemented, there is a possibility that this may be evaluated as a design defect.

(3) Apportionment of Negligence

Even if M's product liability is established, in Case (c), N's worker also made dangerous decisions and took dangerous actions in reliance on the AI's diagnosis. Accordingly, there is a possibility that the amount of damages may be reduced through apportionment of negligence, taking into account N's negligence in failing to take educational and managerial measures to ensure that workers do not blindly follow AI output that may lead to danger, as well as the worker's own negligence.[121]

---

[121] With respect to a worker having performed a dangerous operation in reliance on the AI's output, there is a possibility of apportionment of negligence in light of N's negligence in supervision and management; whether the worker's negligence can be considered as part of the apportionment of negligence in relation to N is also an issue. Where damage has arisen or expanded due to the negligence of an employee, it is generally recognized that the negligence of the employee may be taken into account in calculating the amount of compensation payable to the victim-employer (Yoshimura, Ryoichi (2024). *Fufo Koui Ho (Dai 6-pan) [Tort law (sixth edition)] (in Japanese).* p. 262. Yuhikaku); however, cases are also conceivable in which it was unavoidable for the worker to have relied on the erroneous output of the AI. For a discussion of cases involving autonomous vehicles where the victim acted in expectation of the autonomous driving system applying the brakes, see Sano Makoto

4.2.5 Scope of the Discussion of Hypothetical Case 6

As a supplementary note, the foregoing discussion of product liability and other matters does not extend to all machines and robots that operate through software including AI. In determinations of defect, the characteristics of each individual robot, the intended usage environment, the nature of the User, and other factors constitute important considerations, and for machines and robots other than AMRs, evaluations may differ with respect to, for example, the following factors, and the level and content of safety required may accordingly differ.

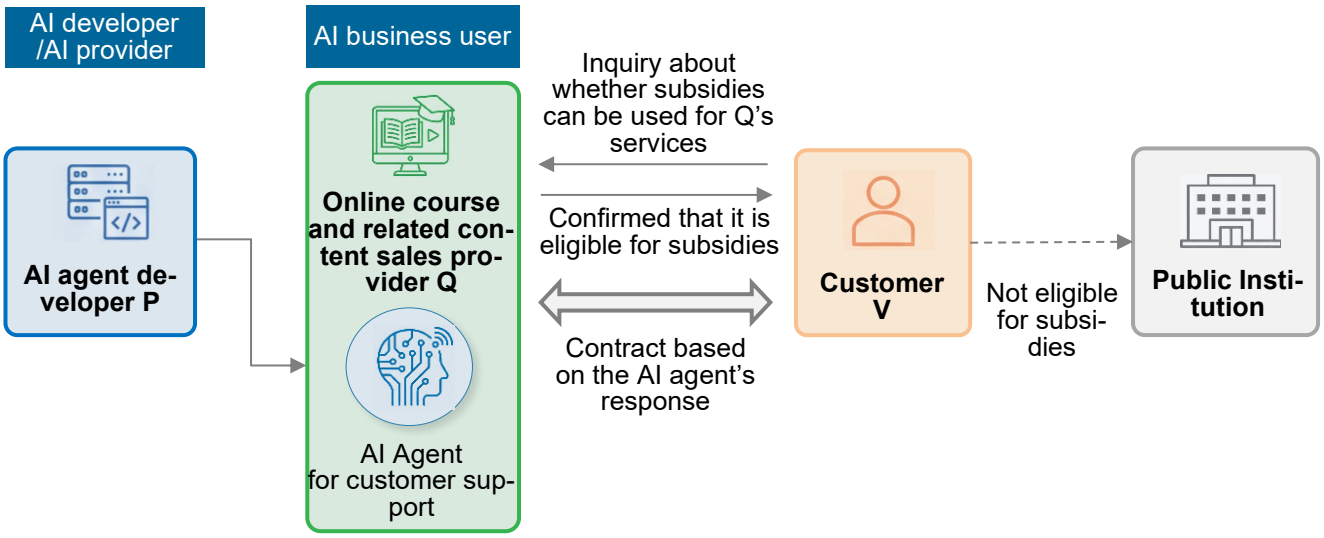
The AMR in Hypothetical Case 6	Comparison with Other Robots
● **Operates only in enclosed spaces such as factories and warehouses** ● **The User is in a position where it should conduct risk assessments under applicable laws** ● **The User is capable of comprehensively managing the area of operation** ● **Collaboration with workers subject to the User's direction and supervision** ● **Does not presuppose operation by humans** ● **No detailed rules or safety standards exist**	● Where the robot operates in more open spaces ⇒ Interaction with children, elderly persons, and the like, as well as unanticipated environmental changes, must also be considered, and a more general level of safety may be required ● Where appropriate maintenance by the User cannot necessarily be expected ⇒ Designs that take into account the User's knowledge, and warnings/instructions that consider comprehensibility, may be necessary ● Where the robot has a function and form similar to machines that have traditionally been operated by humans ⇒ Statistical safety and comparison with the operational ability of a reasonable person may be more readily conducted ● Where regulations or safety standards exist in the relevant domain ⇒ Compliance with regulations and safety standards may constitute an important factor in determinations of defect

(2018). *"Tasutojisya no Sekinin no Hutan no Arikata [The Allocation of Liability Among Multiple Parties."* Fujita Tomoyuki (ed.) *Jidounten to Ho [Autonomous Driving and Law] (in Japanese)*. pp. 211 et seq. Yuhikaku. Furthermore, where a worker's actions contributing to the accident were so unexpected, as with Note 114 above, it is conceivable that the causal relationship between the AMR's defect and the rights infringement may be denied.

4.3 Hypothetical case 7: Supplementary Discussion — AI Agents

Q, which operates a business selling online courses and educational materials, has introduced a so-called AI agent and has automated the majority of its customer support operations. The AI was developed by AI service developer P and is provided to Q as a service; it is composed of multiple AI models, including a multimodal large language model (LLM) that processes voice calls, text, and images with customers. The AI also has the capability, when a customer submits an inquiry, to access Q's internal response manuals and product information as well as information on the internet, collect the necessary information, and reflect that information in its output. As a result of introducing this AI, Q became able to respond rapidly to a significantly greater volume of inquiries than before, the waiting time for customers to receive responses was substantially reduced, and — compared with the conventional automated response system — more specific answers became available even for inquiries submitted on holidays or late at night.

Customer V contacted Q's customer support and asked whether the online course sold by Q was eligible for a public subsidy on tuition fees. Notwithstanding the fact that Q's internal information classified the course as "ineligible for the subsidy," the AI searched for and reviewed information available on the internet indicating that the subsidy was widely applicable, and on that basis provided V with a response and solicitation to the effect that the course "is eligible for the subsidy and should be considered for purchase." V enrolled in the course, but subsequently it came to light that the course was not eligible for the subsidy.



In recent years, the development and deployment of advanced systems referred to as AI agents or agentic AI (hereinafter simply referred to as the "AI agent") — which aim to automate and streamline complex business processes directed toward achieving certain objectives, without presupposing individual data inputs or instructions — has been advancing. With respect to such systems, efforts toward actual deployment and business demonstrations aimed at further

advancement are already progressing, and it has been pointed out that new risks are expanding and emerging compared with conventional AI. However, because risk assessment and governance methods specific to these AI systems are still at a stage of rapid international deliberation[122] and are heavily dependent on individual technologies and use cases, it is at present difficult to set out in specific, case-by-case terms the liability of each party.

For this reason, the Study Group limits its present discussion to setting out the fundamental approach to liability in the utilization of AI; however, these approaches may also serve as a foundation when considering advanced systems such as AI agents. The Hypothetical Case 7 is presented as one example illustrating this point.

Where a company provides a product description that differs from the facts, this may constitute a tort (Article 709 of the Civil Code) as a violation of the duty to explain incidental to the transaction.[123]

Where an AI agent is used in customer support, such AI may qualify legally as either assistive/supportive AI or dependable/substitutive AI, and it is considered that the following points are important in determining the liability of each party.

(1) Liability of Q as the AI Business User

(A) Where the AI Qualifies as Dependable/Substitutive AI

For the AI in question to qualify as dependable/substitutive AI, it is a prerequisite — as set out in 2.2.2 above — that a necessity to entrust judgment to the AI be recognized. As in the hypothetical case above, the significant reduction in customer waiting times and the increased flexibility in handling inquiries on holidays and late at night constitute effects that would be difficult to achieve through human effort alone, and it is considered that these circumstances provide the foundation for a necessity to entrust certain business judgments to the AI.

Further, in terms of the quality and accuracy of the AI, it is necessary that the overall business process — including measures such as introducing human judgment for high-risk consultations or complex responses — performs at a level equal to or higher than that of a reasonable person engaged in the same type of work. In this regard, a heightened duty of care is readily recognized from a consumer protection perspective with respect to product descriptions provided by a business to consumers, particularly those relating to matters affecting the important interests of consumers (see also the judicial precedents cited in note 123 above). In the hypothetical case above, key issues include: whether the AI system was designed to respond accurately to the question of whether the public subsidy applies — a matter that constitutes an important motivation for the consumer's decision to purchase — and, if it was so designed, what unusual sequence of events

---

[122] In Japan, as part of the review process for updating the AI guidelines for business in light of developments in AI agents, the risks associated with AI agents and the governance measures that each party should take as necessary are being examined, as noted in the Ministry of Internal Affairs and Communications document cited above (note 22, p. 4).

[123] See also the Otsu District Court decision of October 3, 2003 (unreported), which recognized a duty to explain in a similar case.

caused the erroneous response in this case; and, where it is difficult for the AI system alone to achieve a sufficient level of response accuracy, whether accuracy across the overall business system was ensured, for example by routing such inquiries to a human operator.

Where the AI qualifies as dependable/substitutive AI under the foregoing analysis, Q's duty of care is converted into a duty relating to the proper construction and operation of the business process incorporating the AI system.

In constructing and operating a business process that incorporates an AI agent, it is a characteristic of AI agents — as illustrated by the hypothetical case above — that they use external information independent of the AI system itself (such as internal databases and external websites) and external tools (such as internal inventory APIs), and autonomously and sequentially select and execute these in order to carry out processing aimed at achieving an objective; consequently, the risks that may arise and the content of measures that an AI business user may take may differ depending on the manner of such use. As noted in note 122 above, the specific content of governance that should be considered in the AI Guidelines for Business and other instruments is expected to be elaborated in the future; however, the following matters, among others, may become issues as to whether they are required. It is considered that, provided that reasonably practicable measures are taken in accordance with the content of the AI Guidelines for Business, the likelihood of negligence being found can be reduced even if the AI ultimately produces an erroneous output.

- Duty of care relating to the construction of a business process incorporating an AI system: It is considered that issues will arise as to whether an AI system satisfying the desired level of accuracy and safety was used, and whether the business process — including human involvement where necessary — was properly constructed.
- Duty of care relating to the operation of the business process: It is considered that issues will arise as to whether measures were required, such as feeding back erroneous responses and improving the system architecture on that basis, and displaying a disclosure that responses are generated by AI (possibly containing errors) [124].

(B) Where the AI Qualifies as Assistive/Supportive AI

Unlike in (A) above, where the AI qualifies as assistive/supportive AI — for reasons such as insufficient response accuracy — Q should use the AI agent solely as an aid or support for its own judgment, and as noted above, a heightened duty of care is readily recognized from a consumer protection perspective with respect to explanations provided to customers in such settings. Where an erroneous AI output is used as a response to a customer regarding a matter that is material to

[124] The prior disclosure to consumers, in an intelligible manner, that responses are generated by AI rather than by a human and that there is a possibility of erroneous responses, may be an important factor to decide that reasonably practicable measures have been taken.

the customer's decision to purchase, there may be a finding of a violation of a duty of care with respect to the judgment of the employee who relied upon that response (in which case Q would bear employer liability), and there may also be a finding of a violation of a duty of care with respect to an insufficiency in the staffing arrangements for verifying the content of the AI's output.

(2) Liability of P as the AI Developer and Provider

As with (1) above, the governance required of AI developers and providers is expected to be elaborated in the future in the AI Guidelines for Business and other instruments; however, taking into account the characteristics and risks — namely, the autonomous and sequential selection and execution of external information and external tools as described above — the key issue is whether appropriate design measures and clear explanations were provided. The standard of these obligations differs depending on whether the AI is provided as dependable/substitutive AI or as assistive/supportive AI.

(A) Where the AI Qualifies as Dependable/Substitutive AI

Where the AI qualifies as dependable/substitutive AI, it is considered that important issues include: whether P took necessary design measures — including ensuring that the AI handles external information and tools appropriately — in order to achieve and maintain the anticipated level of accuracy in substituting for part of the judgment and actions of human operators; and whether P considered important factors for controlling risks including the scope within which human involvement should be interposed and provided Q with explanations of these factors.

(B) Where the AI Qualifies as Assistive/Supportive AI

On the other hand, where the AI qualifies as assistive/supportive AI, the central issue is whether P provided explanations regarding the AI's functions, settings for use, methods of use, and important risks — on the premise that Q would verify and correct the appropriateness of the final output. In such cases, where significant risks arise from the use of external information and tools of the kind described above, clear explanations may be required. Furthermore, where the AI agent shows a tendency to exhibit unexpected behavior and risks arise that are difficult for the AI business user to address appropriately, design measures to mitigate such risks may be required.

Chapter 5 Issues Relating to Evidence and Procedure

5.1 Issues Relating to Evidence

In the preceding chapters, the approach to the interpretation and application of substantive law with respect to the determination of liability when rights are infringed or damages are suffered has been set out. However, in resolving actual cases, specialized knowledge relating to risk control in light of the technical characteristics of AI is required, and in many instances it is considered that the relevant materials are held exclusively by AI developers, providers, and business users. When a person who has suffered damage exercises rights in court as a plaintiff, the burden of pleading and proof in respect of the cause of action under tort liability or the Product Liability Act rests with that plaintiff; however, in light of the technical specialization and the unequal distribution of evidence described above, difficulties are expected in pleading and proving, in particular, elements such as negligence (Article 709 of the Civil Code) and defects (Article 3 of the Product Liability Act).

With respect to these difficulties in pleading and proof, mechanisms for evidence disclosure have been utilized in the past; and in technically specialized areas such as medical litigation and environmental litigation in particular, various doctrines relating to the prima facie presumption of negligence or defects (effectively shifting the burden of proof) have been applied under certain conditions. The Study Group, as a matter incidental to its consideration of substantive law issues, also discussed the applicability of these existing mechanisms and doctrines to AI cases, as well as future challenges.

5.1.1 Orders to Produce Documents

An order to produce documents (paragraph (1) of Article 223 of the Code of Civil Procedure) is a mechanism whereby a court orders the holder of a document to produce it where that holder bears a duty to produce documents (Article 220 of the Code of Civil Procedure). Where an order to produce documents has been issued but a party to the litigation fails to comply with the order, or destroys the document with the intent to obstruct the adverse party's use of it, the court may treat the adverse party's allegations regarding the content of that document as true (paragraphs (1) and (2) of Article 224 of the Code of Civil Procedure); furthermore, where it is extremely difficult for the party bearing the burden of proof to prove the facts to be proven by other evidence, the

court may treat the adverse party's allegations regarding the facts to be proven as true (paragraph (3) of the same Article).[125]

As noted above, where the holder of a document is found to bear a duty to produce that document, an order to produce documents will be issued; accordingly, whether AI developers, providers, and business users bear such a production duty is an important issue. Under the Code of Civil Procedure, the holder of a document bears, in addition to specific production duties under items (i) through (iii) of Article 220 of the same Code, a general duty to produce documents unless an exception under item (iv) of the same Article applies.

In this regard, documents created in the course of the development, provision, and use of AI may contain certain trade secrets, and it is anticipated that the exemption from the production duty based on technical or professional secrets — one of the exceptions — (item (iii) of paragraph (1) of Article 197 of the Code of Civil Procedure and item (c) of item (iv) of Article 220 of the Code of Civil Procedure) will become a contested issue.[126] Whether information constitutes a trade secret is to be determined through a balancing of various circumstances, including the nature and extent of the disadvantage to the holder if the information is disclosed, against the necessity of the document as evidence.[127]

The Study Group noted that if decisions not to disclose evidence become more frequent because parties prioritize monetary settlement over the disadvantages of disclosure, the collection of information necessary to enhance the safety of AI becomes difficult; and it was pointed out that the design of mechanisms to ensure AI safety across society as a whole is important, including with a view to coordination with mechanisms outside civil litigation, such as mechanisms for the collection of information for regulatory and supervisory, while also considering the balance with trade secrets and other interests, purposes by administrative authorities.

---

[125] Note that where an order to produce documents is sought against a third party who is not a party to the litigation, in addition to the ordinary procedure, a hearing of that third party (paragraph (2) of Article 223 of the Code of Civil Procedure) is required; and there are also differences in procedure and effect — for example, the effect of non-compliance with the order is a non-criminal fine (administrative fine) rather than a legal fiction of truth (paragraph (1) of Article 225 of the Code of Civil Procedure).

[126] Note that where a document is determined — based on its purpose of creation, its content, the circumstances in which it came to be held by the current holder, and other circumstances — to have been created solely for the use of internal persons and not intended for disclosure to external persons, and where there is a risk that disclosure of the document would cause non-negligible disadvantage to the holder — such as an infringement of an individual's privacy or an impediment to the free formation of the will of an individual or organization — the document may in some cases be exempt from the production duty as a self-use document (item (d) of item (iv) of Article 220 of the Code of Civil Procedure). However, with respect to documents created in the course of the development, provision, and use of AI, it is considered that the exemption from the production duty based on technical or professional secrets, as described in the main text, is typically more likely to become a contested issue.

[127] The Supreme Court Decision of November 25, 2008 (Minshu Vol. 62, No. 10, p. 2507) held that the right to refuse production of a document under item (c) of item (iv) of Article 220 and item (iii) of paragraph (1) of Article 197 of the Code of Civil Procedure is limited to cases in which the professional secret described in the document in question constitutes a secret worthy of protection, and that whether the information is a secret worthy of protection should be determined through a balancing of various circumstances, including the nature and extent of the disadvantage to the holder if the information is disclosed, against the nature and content of the civil case and the degree to which the document is necessary as evidence in that civil case.

5.1.2 The Doctrine of de facto Presumption of Negligence

With respect to the finding of negligence, particularly in areas such as medical litigation and environmental litigation, the de facto presumption of negligence has been utilized to alleviate the plaintiff's burden of pleading and proof under certain conditions. There are diverse categories of the de facto presumption of negligence, including for example the following.

- **Presumption of negligence premised on a breach of a duty of conduct**: The Supreme Court Judgment of January 23, 1996 (Minshu Vol. 50, No. 1, p. 1) held that, where a physician fails to follow the precautions for use stated in a package insert when administering a pharmaceutical and a medical accident occurs as a result, the negligence of that physician is presumed unless there is a particular rational reason for not following those precautions.

- **Alternative finding of negligence premised on maxims of experience**: The Supreme Court Judgment of May 10, 1957 (Minshu Vol. 11, No. 5, p. 715) alternatively found either negligence consisting of defective injectable solution or negligence consisting of inadequate disinfection of the injection instrument, in respect of suppuration caused by a physician's injection. Furthermore, the Supreme Court Judgment of July 28, 1964 (Minshu Vol. 18, No. 6, p. 1241) alternatively found negligence consisting of inadequate disinfection of either the injection instruments, the hands of the practitioner, or the injection site on the patient, in respect of suppuration caused by a physician's injection.

- **Comprehensive presumption of negligence premised on uncertainty :** The Gifu District Court Judgment of July 20, 1994 (Hanrei Times No. 861, p. 49), in an action seeking an injunction against the construction of a tidal barrage on the Nagara River, held that a rational judgment based on scientific and technical expertise must be made regarding the safety of the barrage, and that, taking into account the fact that materials relating to the safety of the barrage in question were held by the defendant, the defendant must first prove, on the basis of adequate grounds and materials, that there are no deficiencies in safety, from the perspective of fairness.

The de facto presumption of negligence is a doctrine that should be applied only exceptionally under the principle that the plaintiff bears the burden of pleading and proof, and requires individual, case-specific examination in each case; accordingly, the approach set out in the above judicial precedents cannot be applied to AI cases as-is.
Nevertheless, features such as technical specialization and the unequal distribution of evidence may also be applicable to AI-related cases. In addition to this necessity, it is considered that, where exceptional circumstances exist to warrant modification of the general principle that the plaintiff bears the burden of proof —taking into seriousness of legal interests such as life and body, violations of certain duties to act, empirical rules, or a high degree of uncertainty — there

may be cases in which a de facto presumption should be considered, with reference to the approach set out in the above judicial precedents.

5.1.3 The Doctrine of de facto Presumption of Defects

In actions for damages based on product liability as well, judicial precedents have utilized the factual presumption of defects as a doctrine for alleviating the burden of pleading and proof borne by the plaintiff with respect to "defects" in a product.

Specifically, the factual presumption of defects has been utilized whereby it is sufficient for the plaintiff who has suffered damage to plead and prove that: (A) the product was being used in accordance with its ordinary method of use, yet (B) an abnormality causing damage to person or property occurred (Sendai High Court Judgment of April 22, 2010 [unreported]; Tokyo High Court Judgment of February 13, 2013, Hanrei Jiho No. 2208, p. 46).

It is considered that the application of this doctrine is also possible with respect to machines and robots that operate autonomously, as discussed in 4.2 and Hypothetical Case 6 above. However, compared with conventional products intended to be used by humans, cases are anticipated in which it is difficult to determine when requirements (A) and (B) above are satisfied in relation to products that operate autonomously without the intervention of human judgment or action.

(1) Content of "Ordinary Method of Use"

Even where a machine autonomously determines and acts upon individual behaviors, it is considered that it is possible to evaluate whether those behaviors are consistent with the "ordinary method of use" by examining the operational aspect of the machine. For example, a state in which the machine does not deviate in material respects from the operating environment and methods of use specified by the manufacturer may qualify as the "ordinary method of use."

(2) Content of the "Abnormality" Causing Damage to Person or Property

On the other hand, the greater the variation in the behavior of a machine, the more difficult it may become to determine whether a behavior is the result of the intended design or is "abnormal" behavior.[128] For example, where a machine or robot suddenly begins to engage in dangerous movement without rational reason, this is readily evaluated as "abnormal"; whereas in cases where it is unclear whether or not behavior is "abnormal" — such as where various environmental factors may have contributed to the dangerous behavior — it is also conceivable that the victim would be required to prove that the behavior deviated from the intended design.

[128] Regarding autonomous vehicles, for the observation that even if used in accordance with the ordinary method of use, an automated driving system cannot prevent all accidents, and accordingly the issue becomes whether the accident was one that the automated driving system was required to prevent — and that proof of this point is necessary — see Tomoyuki, Fujita (2018). "Jidounten wo Meguru Minjisekinin Hosei no Shoraizo [The Future of Civil Liability Law Relating to Autonomous Driving]" (in Japanese) Tomoyuki Fujita (ed.) *Jidounten to Ho [Autonomous Driving and Law]* . p. 279 Yuhikaku.

The Study Group noted that, even where a machine or robot that operates autonomously is operated in accordance with its ordinary method of use, the actual behaviors it exhibits depend on the behavior of the software such as the AI, and that determining the abnormality of its operation may therefore become difficult in some cases; accordingly, it was pointed out that a doctrine capable of directly addressing technical and scientific complexity is required.

5.1.4 Determination of Causal Relationships

Furthermore, under both general tort liability and product liability, a causal relationship between negligence or defects and the damage is a requirement (as set out in 2.1.1 and 2.1.2 above); however, where the risk of damage arising from AI output can only be understood in statistical terms, it is anticipated that the determination of whether or not a causal relationship exists may become difficult.

For example, consider a setting involving a dependable/substitutive AI visual inspection system of the kind described in Hypothetical Case 5, where it is assumed that the detection accuracy of the AI that the AI business user should have used was 90%, but the AI actually used had an accuracy of 70%, and the AI failed to detect a foreign object that a reasonable person would not have overlooked. In such a case, it may be difficult to assess whether using the AI at the requisite standard would have prevented the injury suffered by the individual plaintiff, or whether the outcome would have occurred regardless.[129]

(1) Resolution Through Factual Findings

The proof of causal relationships is said to require not proof to the standard of natural-scientific certainty that admits of no doubt whatsoever, but proof of "a high degree of probability" assessed in light of rules of experience.[130] Even where the magnitude of a statistical risk does not in itself suffice to establish a high degree of probability, it may nonetheless constitute one of the circumstances supporting such a finding; accordingly, a high degree of probability may be established by considering such statistical evidence together with other circumstantial facts[131] (see also the judicial precedents cited in note 132 below).

(2) Resolution Through Interpretive Approaches

[129] Note also, as a point raised in the discussion in the United States, that in a hypothetical case where an autonomous vehicle manufacturer fails to conduct the necessary volume of pre-market testing and sells vehicles with a collision probability equal to 3 times in 10,000 trips — and where, due to a defect in the design of the operating system, on average one extra collision occurs for every 10,000 trips — it has been pointed out that it is difficult to determine which accidents could have been avoided if the operating system had not contained a design defect (see note 105 above, Mark A. Geistfeld, p. 204).

[130] See Supreme Court Judgment of October 24, 1975, *Minshu Vol. 29, No. 9,* p. 1417 (the Lumbar Puncture Case).

[131] The Supreme Court Judgment of February 6, 1969, *Minshu Vol. 23, No. 2,* p. 195 recognized a causal relationship between X-ray irradiation and skin cancer, based on: the existence of a statistical correlation between irradiation and the development of cancer (in particular skin cancer); the timing, quantity, number, and location of the irradiation in the case in question; and the fact that the victim's skin cancer developed only at the irradiated sites, among other facts.

Where resolution through the factual findings described in (1) above is difficult, certain interpretive approaches have been adopted in past judicial precedents, which may also serve as a reference in the AI context.

First, some judicial precedents concerning Minamata disease adopted an approach of recognizing damage amounts in proportion to probability; in relation to certain plaintiffs for whom a high degree of probability of a causal relationship could not be established, the amount of damages was determined by multiplying the damage suffered by the plaintiffs by the probability that they had contracted Minamata disease.[132] These holdings presuppose pollution cases and medical limitations, and it is considered that their scope should be interpreted as limited; moreover, because there is also the challenge of how to reasonably estimate the probability that damage is attributable to AI output, it is considered that the scope of their application is limited.[133]

In addition, in medical malpractice cases, there is an approach that, in cases where a high degree of probability of a causal relationship with the ultimately resulting serious outcome cannot be established, the infringement of a reasonably substantial possibility of survival or of an expectation interest in receiving appropriate medical treatment is recognized, and an award of solatium (consolation damages) is granted. However, it is considered that this approach also presupposes serious infringement of legally protected interests such as life and physical integrity, and is appropriate only where it is possible to identify a suitable alternative legally protected interest; accordingly, its application should be considered with caution.

5.2 Procedural Issues Relating to International Disputes

Cases in which the development, provision, and use of AI services cross national borders are increasing; and associated with this, situations are anticipated in which a person who has suffered damage from AI output considers pursuing a claim against an overseas business operator, or in which a Japanese AI developer or similar party receives a claim from a person residing abroad. Where an attempt is made to resolve such cross-border disputes through court proceedings,[134]

---

[132] The Tokyo District Court Judgment of February 7, 1992 (Hanrei Jiho, April 25, 1992 issue, p. 3) held that, given that among those whose applications for recognition of Minamata disease had been rejected, no small number — 19 out of 71 cases (27%) — were posthumously diagnosed pathologically as having had Minamata disease, requiring a high degree of probability in the proof of causal relationships would unilaterally impose upon the plaintiff the burden arising from the limitations of medicine, which would be contrary to the equitable allocation of damage; accordingly, where a "reasonably substantial possibility" of having had Minamata disease is recognized, it is considered more appropriate to reflect the degree of that possibility in the calculation of the amount of damages (solatium (consolation damages)). Furthermore, the Osaka District Court Judgment of July 11, 1994, *Hanrei Jiho No. 1506,* p. 5 recognized a "high degree of probability" with respect to plaintiffs who met the administrative recognition criteria for Minamata disease, while adopting a proportional allocation method with respect to plaintiffs who did not meet the recognition criteria — dividing the probability of having had Minamata disease into four levels of "40%, 30%, 20%, and 15%" according to the combination of symptoms, and multiplying a base amount by that proportion.

[133] While the Study Group also received observations that the scope of these approaches is limited, it was also noted that — at least in cases where there is no clue whatsoever other than the statistical tendencies of AI output and risk can only be evaluated in probabilistic terms — it is not entirely precluded to consider recognizing a damage amount in proportion to probability.

[134] In disputes in which the design and operation of AI systems and their behavior are at issue, a high degree of technical expertise is often required for fact-finding and the evaluation of liability. In such cases, the utilization of alternative dispute

the key issues are: which country's courts have jurisdiction over the dispute in question (international jurisdiction); which country's laws are to be applied (governing law); and whether a judgment obtained in one country can be enforced in another country (enforcement of foreign judgments).

The approach to the interpretation and application of these issues is discussed in the Guidelines on Electronic Commerce and Information Property Transactions (the Ministry of Economy, Trade and Industry, February 2025 edition) (hereinafter "the Guidelines"), and it is considered that the content thereof may be applied by analogy to disputes arising from the development, provision, and use of AI; accordingly, an overview is set out below.[135]

For example, where a plaintiff located in Japan is considering a claim against an overseas AI business operator, the plaintiff would determine in which country to pursue the claim on the merits and the enforcement proceedings, taking into comprehensive account the substantive and procedural laws of each country.

The approach in cases where proceedings are conducted in Japanese courts against a business operator located abroad may be organized as described in 5.2.1 through 5.2.3 below.[136]

5.2.1 International Jurisdiction

International jurisdiction concerns the question of whether Japanese courts have jurisdiction over the dispute in question; where jurisdiction does not exist, even if an action is filed in a Japanese court, it will in principle be dismissed.

First, where an arbitration agreement exists, a lawsuit filed in a Japanese court with respect to a dispute falling within the scope of that agreement will be dismissed upon the application of the defendant (paragraph (1) of Article 14 of the Arbitration Act).[137]

Where no arbitration agreement exists, or where one has not been validly concluded, the issue is whether the case falls within the categories enumerated in the Code of Civil Procedure as cases in which Japanese courts are recognized to have jurisdiction when an action is filed in a Japanese court.[138]

---

resolution (ADR) procedures such as arbitration and mediation involving persons with specialized expertise in AI technology, in addition to resolution through court proceedings, may in some instances lead to a resolution that is better suited to the actual circumstances of the dispute.

[135] The following organizes the principal issues with claims based on tort liability — the subject of this document's examination — in mind; however, in relation to individual cases, more detailed special provisions and other instruments may be relevant. For details, please refer to the Guidelines, p. 336 et seq.

[136] Note that where the plaintiff is a consumer, the effectiveness of provisions designating foreign courts or foreign law may be limited by consumer protection provisions such as Article 11 of the Act on General Rules for Application of Laws.

[137] See the Guidelines, p. 352.

[138] See the Guidelines, p. 352.

- **Jurisdiction based on the defendant's domicile and similar factors** : Where the defendant is a corporation, Japanese courts are recognized to have jurisdiction if the defendant's principal office or place of business is located in Japan (paragraph (3) of Article 3-2 of the Code of Civil Procedure).
- **Jurisdiction based on the type of action** : There are various types of jurisdiction based on the type of action; the following are the types of particular importance in relation to claims based on tort liability.[139] First, with respect to overseas AI business operators that do not have a base of operations in Japan, jurisdiction over "persons who conduct business in Japan" is considered (item (v) of Article 3-3 of the Code of Civil Procedure). In determining whether or not such a business operator conducts business in Japan, the fact that the business operator provides services in the Japanese language is an important consideration. Furthermore, in relation to claims based on tort liability, jurisdiction based on "the place where the tort occurred" is considered (item (viii) of Article 3-3 of the Code of Civil Procedure). The place of the tort includes not only the place of the tortious act but also the place where the result occurred, and where the result of the rights infringement has occurred in Japan, jurisdiction is recognized under that item. [140]
- **Jurisdiction by agreement** : Since international jurisdiction may be determined by agreement of the parties (paragraph (1) of Article 3-7 of the Code of Civil Procedure), where there is an agreement designating Japanese courts as the courts of jurisdiction, the international jurisdiction of Japanese courts is recognized.

5.2.2 Governing Law

The next issue relating to governing law is which country's laws are to be applied to the dispute in question.

Where a case is heard in a Japanese court and governing law is to be determined, the court will first apply the law of the country selected by agreement of the parties — pursuant to Article 7 of the Act on General Rules for Application of Laws (hereinafter "the Act on General Rules") — if the parties had reached an agreement at the time of the transaction regarding which country's law would govern the contract.[141] Where no such agreement exists, the court proceeds to consider the governing law in accordance with other provisions of the Act on General Rules. The following categories are of particular importance in relation to claims based on tort liability.

---

[139] See the Guidelines, pp. 364–365.

[140] This is subject to the exception where the occurrence of the result of a tortious act committed abroad within Japan was one that could not normally be foreseen (item (viii) of Article 3-3 of the Code of Civil Procedure; see the Guidelines, p. 364). Note also that, where foreign law is to be applied with respect to a tort, if the facts to which that foreign law should be applied do not constitute a tort under Japanese law, a claim for damages or other relief based on that foreign law cannot be made, and the victim may claim only such damages or other relief as is recognized under Japanese law (See Article 22 of the Act on General Rules and p. 364 of the Guidelines.).

[141] See the Guidelines, p. 345.

- The general rule governing the governing law applicable to tort law is Article 17 of the Act on General Rules, and in principle "the law of the place where the result of the tortious act occurred" applies; where the result of the rights infringement occurred in Japan, Japanese law becomes the governing law.[142]
- In cases of defamation or damage to reputation, the law of the place of habitual residence of the victim (or, where the victim is a corporation or other association or foundation, the law of the place where its principal place of business is located) becomes the governing law (Article 19 of the Act on General Rules). Although views differ as to whether Article 19 of the Act on General Rules or Article 17 (or Article 20) of the Act on General Rules applies depending on the nature of the rights infringement, where a plaintiff located in Japan pursues a claim against an overseas AI business operator, it is considered that either the law of the place of habitual residence of the victim or Japanese law — as the law of the place where the result occurred in Japan — will apply, and the conclusion is that Japanese law is likely to be the governing law.[143]

Where Japanese law is applied pursuant to the foregoing rules, the approaches set out in this document are applicable in relation to claims based on tort liability

5.2.3 Recognition and Enforcement of Foreign Judgments and Foreign Arbitral Awards

Where a judgment or other ruling granting the claim is ultimately obtained, the enforcement of that judgment will be considered; however, where the claim was pursued in a foreign court or through foreign arbitration proceedings, the issue is whether the resulting judgment or award can be enforced in Japan.[144]

- A party seeking to enforce in Japan a favorable judgment rendered by a foreign court must file an "action for enforcement of judgment" in a Japanese court. Where the foreign judgment is a final and conclusive judgment, and the requirements are satisfied — including that the content of the judgment and the litigation procedure do not violate public policy and good morals in Japan — the judgment may be enforced on the basis of the foreign judgment (Article 118 of the Code of Civil Procedure).
- With respect to awards rendered in foreign arbitration proceedings as well, there are cases in which enforcement in Japan is possible pursuant to the New York Convention or the Arbitration Act, and a party seeking enforcement must file a "motion for enforcement

[142] See the Guidelines, p. 360.
[143] See the Guidelines, pp. 365–366.
[144] See the Guidelines, p. 377 et seq.

decision." Where requirements such as the validity of the arbitration agreement and the absence of a violation of public policy and good morals are recognized, enforcement may be effected on the basis of the foreign arbitral award.

Guidance on the Interpretation and Application of Civil Liability in the Utilization and Application of AI

［Version 1.0］

April 2026

Digital Economy Division, Commerce and information Policy Bureau, Ministry of Economy, Trade and Industry

AI Industry Strategy Office, IT Industry Division, Commerce and information Policy Bureau, Ministry of Economy, Trade and Industry

The Study Group on Civil Liability in the Utilization of AI
OTSUKA Tadashi (Chair)
INATANI Tatsuhiko
EMA Arisa
KAKINUMA Taichi
SHISHIDO George
SHIRAISHI Tomoyuki
NAKAHARA Taro
HASHIMOTO Yoshiyuki
FUKUOKA Shinnosuke
MATSUO Yutaka