

'General Understanding' on AI and Copyright in Japan*

Tatsuhiko Ueno**

Abstract

This article examines Japan's "General Understanding on AI and Copyright," published on 15 March 2024 following intensive deliberations by the Legal Subcommittee of the Cultural Council established by the Agency for Cultural Affairs. The document addresses three core copyright issues surrounding AI: (i) whether using copyrighted works for AI training infringes copyright, (ii) whether AI-generated content infringes copyright, and (iii) whether AI-generated content qualifies for copyright protection.

Regarding the training stage, the article analyzes Japan's 'broad' TDM (text and data mining) exception under Article 30-4 of the Copyright Act, introduced in 2009 — making Japan the first country worldwide to establish such provisions. The General Understanding clarifies that this exception does not apply when an "enjoyment purpose" coexists with analytical purposes, particularly when training intentionally aims to reproduce original expressions. The proviso excluding acts that "unreasonably prejudice the interests of the copyright owner" is examined through two contested scenarios: targeted training on specific creators' works, and the use of database works created for TDM activities.

For the output stage, this article discusses infringement standards requiring both similarity and dependence. The General Understanding establishes that dependence is generally presumed when AI has learned from copyrighted works, even if users were unaware of them. However, it suggests dependence might be denied where technical safeguards ensure trained works' expressions cannot be generated.

Written by a Subcommittee member who helped formulate the General Understanding, this article clarifies interpretations of Japan's latest position on AI and copyright issues while acknowledging the need for ongoing revision as technology and international discussions evolve.

Keywords

AI training, Machine learning, Copyright, Japan, TDM, Text and data mining

* This article is the translation of Tatsuhiko Ueno, 'AI と著作権に関する考え方について' [General Understanding on AI and Copyright] (2024) 215 Hō no Shihai 87 (in Japanese).

** Professor at Waseda University (Tokyo); Director of the Research Center for the Legal System of Intellectual Property (RCLIP); Director of the Institute of Comparative Law.

I Introduction

Legal issues surrounding AI (artificial intelligence) are wide-ranging, and copyright is among the areas with numerous points at issue. Specifically, the key issues include: (i) whether the use of another person's copyrighted works, etc., for AI training constitutes an infringement of copyright, etc.; (ii) whether the content generated by generative AI constitutes an infringement of copyright, etc.; and (iii) whether the content generated by generative AI is eligible for copyright protection.¹

Debates on these points at issue have intensified following the rapid development of generative AI technology starting around 2023. In particular, since it is well-known that the Japanese Copyright Act contains a broad TDM copyright exception (Article 30-4(ii)), there have been heated discussions in Japan on AI and copyright.

Accordingly, in Japan, the Intellectual Property Strategy Headquarters' "Intellectual Property Strategic Program 2023" stated that "it is desirable to clarify perspectives on various instances," including "Perspectives on the creative contribution of users required for AI-generated content to be recognized as copyrighted work," "Perspectives on copyright infringement when AI-generated content is similar to the originals," and "Perspectives on the clause 'unreasonably prejudice the interests of the copyright owner' as defined in Article 30-4 of the Copyright Act when using copyrighted works to develop AI (trained models)".² In response, the Legal Subcommittee under the Copyright Subdivision of the Cultural Council (hereinafter referred to as the "Subcommittee") held meetings from the first session (July 26, 2023) to the seventh session (Feb. 29, 2024),³ and then published as of March 15 of the same year, the "General Understanding on AI and Copyright" (hereinafter referred to as the "General Understanding").⁴

¹ See Tatsuhiro Ueno & Koji Okumura (eds), *AI と著作権 [AI and Copyright]* (Keisō Shobō 2024) (in Japanese); Tatsuhiro Ueno, '人工知能と機械学習をめぐる著作権法上の課題' [Copyright Law Issues Surrounding Artificial Intelligence and Machine Learning], in *知的財産紛争の最前線(3) [Frontiers of Intellectual Property Disputes (3)]* (Minjihō Kenkyūkai 2017) 92 (in Japanese); Id. '人工知能と機械学習をめぐる著作権法上の課題——日本とヨーロッパにおける近時の動向' [Copyright Law Issues Surrounding Artificial Intelligence and Machine Learning: Recent Developments in Japan and Europe] (2019) 91-8 Hōritsu Jihō 33 (in Japanese).

² Intellectual Property Strategy Headquarters, 'Intellectual Property Strategic Program 2023 - The transformation towards a society where motivated individuals and players can fully utilize Intellectual Property', 36-37 (June 9, 2023), available at https://www.kantei.go.jp/jp/singi/titeki2/kettei/chizaikeikaku2023_e.pdf.

³ In addition, the Subcommittee also solicited opinions among its members outside of formal meetings regarding the first revised version (dated December 27, 2023) following the fifth session (December 20, 2023) from December 27 to January 4, and regarding the second revised version (dated January 15, 2024) following the sixth session (January 15, 2024) from January 16 to January 17 (in Japanese). In addition, as many as 24,938 opinions were received (73 of which were made by organizations) during Public Comment from January 23 to February 12 (<https://www.bunka.go.jp/seisaku/chosakuken/aiandcopyright.html>).

⁴ Legal Subcommittee under the Copyright Subdivision of the Cultural Council, 'AI と著作権に関する考え方について' [General Understanding on AI and Copyright] (March 15, 2024) (in Japanese), available at

This “General Understanding,” developed through intense debate both within and outside the Subcommittee, comprehensively addresses all relevant issues. It represents a significant achievement in that it demonstrates the culmination of discussions while adhering to the fundamental principles and established interpretations of the Copyright Act. Consequently, it has garnered extensive attention from various quarters and attracted considerable international interest.

On the other hand, the “General Understanding” includes sections that take fairly complex cases into consideration, and its detailed and intricate writing makes certain aspects difficult to understand precisely. In particular, when the content of the “General Understanding” is presented in the media, etc., it inevitably tends to be simplified, which raises concerns that it may lead to misunderstandings among the general public.

Accordingly, this article, written by this author, who participated as a member of the Subcommittee in formulating the “General Understanding,” clarifies how to interpret the “General Understanding,” focusing in particular on the highly contested issues concerning infringement of copyright, etc. (Items (i) and (ii) above), for the purpose of introducing the latest discussions on AI and copyright in Japan.

II Copyright Issues in the Training Stage

1 Identification of Issues

First, there is the issue of whether the use of another person’s copyrighted works, etc. for AI training constitutes an infringement of copyright, etc.

It is generally common and effective to use large volumes of data for training AI. However, when such data includes works, performances, phonograms, broadcasts, or cable broadcasts for which others hold copyrights or neighboring rights (hereinafter referred to as “copyright, etc.” and such materials referred to as “copyrighted works, etc.”), the act of inputting them into a computer may give rise to issues of copyright, etc. This is because, even for the purpose of machine learning, inputting works into a computer constitutes reproduction to the extent that it can be considered to be physical replication (Article 2(1)(xv) of the Japanese Copyright Act⁵) and is therefore subject to the right of reproduction (Article 21 of the Japanese Copyright Act), which means that, even if conducted internally within a company, inputting works could potentially constitute copyright infringement.

https://www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/pdf/94037901_01.pdf.

⁵ Non-official translations of the Japanese Copyright Act are available at <https://www.japaneselawtranslation.go.jp/en/laws/view/3379> and <https://www.cric.or.jp/english/clj/>. Regarding the outline written in English of the Japanese Copyright Act and major cases, see Tatsuhiro Ueno, ‘Chapter 22 (Japan)’ in Silke von Lewinski (ed), *Copyright Throughout the World* (Thomson / West, loose-leaf from 2008); Teruo Doi / Tatsuhiro Ueno, ‘Chapter JAPAN’ in Lionel Bently (ed), *International Copyright Law and Practice* (Matthew Bender/LexisNexis, loose-leaf from 2015).

2 TDM Copyright Exception

However, the Japanese Copyright Act contains a provision limiting rights that permits the use of works for the purpose of “information analysis” or text and data mining (TDM) (hereinafter referred to as the “TDM exception”).

In other words, Article 30-4 of the Japanese Copyright Act provides that “[i]t is permissible to exploit a work, in any way and to the extent considered necessary, in any of the following cases, or in any other case in which it is not a person's purpose to personally enjoy or cause another person to enjoy the thoughts or sentiments expressed in that work; provided, however, that this does not apply if the action would unreasonably prejudice the interests of the copyright owner in light of the nature or purpose of the work or the circumstances of its exploitation: ...,” and Art 30-4(ii) states, “if it is done for use in data analysis (meaning the extraction, comparison, classification, or other statistical analysis of the constituent language, sounds, images, or other elemental data from a large number of works or a large volume of other such data).”

Such TDM exceptions have, in recent years, increasingly been introduced outside Japan as well, i.e., in the United Kingdom (2014), Germany (2017), France (2018), the EU Directive (2019), Switzerland (2019), and Singapore (2021). However, the Japanese TDM exception was originally introduced by the 2009 amendment to the Japanese Copyright Act (Act No. 53 of 2009)⁶ (Article 47-7 of the former Copyright Act amended in 2018 [Act No. 30 of 2018] [Reproduction for Information Analysis, etc.]). In other words, Japan was the first country in the world to introduce a TDM exception.

While rigorous international comparisons of such TDM exceptions are not always straightforward,⁷ the Japanese TDM exception can be considered more robust than those in other jurisdictions, because there are no restrictions on the entities conducting the analysis or its purpose, rights holders cannot opt out of any TDM activities, there are no limitations on the content of permissible acts, and there are no lawful access or lawful copying requirements.⁸ This author has referred to Japan, which was the first country in the world to introduce such a broad TDM exception, as a “Paradise for Machine learning,” and has been proposing the leveraging of the Japanese TDM exception since 2016.⁹

⁶ Regarding the amendment in 2009 see Copyright Division, ‘Commissioner's Secretariat, Agency for Cultural Affairs, Explanation of the revision of the copyright act in 2009’, 03/2010 Journal of the Japanese Group of AIPPI 81.

⁷ For details, see Ueno & Okumura, eds., supra note (1), at 70 et seq. [Ueno].

⁸ For details, see Ueno & Okumura, eds., supra note (1), at 65 et seq. [Ueno].

⁹ See Tatsuhiro Ueno, ‘機械学習パラダイス’ [Machine-Learning Paradise], RCLIP Column (September 9, 2017) (in Japanese), available at <https://www.rclip.jp/jp/publications/2017/09/09/201708column>; Id. ‘著作権法改正が拓く日本の“機械学習パラダイス”’ [Japan's ‘Machine-Learning Paradise’ pioneered by the amendment of the Copyright Act], (2018) 19-2 Business Hōmu 1 (in Japanese); Id. ‘情報解析と著作権——『機械学習パラダイス』としての日本’ [TDM and Copyright: Japan as a ‘Machine-Learning Paradise’] (2021) 36-6 Jinkōchinō 745 (in Japanese). It should be noted, however, that the training stage (input) and the generation stage (output) must be distinguished, and that even if the Japanese TDM exception is described as a ‘Paradise for Machine

3 “General Understanding”

However, since around 2023, as certain types of generative AI achieved rapid development, renewed attention has been focused on this TDM exception. Therefore, the interpretation of the Japanese TDM exception became a major issue in the Sub-Committee as well, especially with regard to the following two points.

(1) Coexistence type for the purpose of ‘enjoyment’

Article 30-4(ii) of the Japanese Copyright Act cites “information analysis” (TDM) as an example of cases in which “it is not a person's purpose to personally enjoy or cause another person to enjoy the thoughts or sentiments expressed in that work” (hereinafter referred to as “non-enjoyment use purposes”) as stated in the main text of the introductory clause of the same article.¹⁰

On this point, the “General Understanding” states that “even if a certain use is carried out for non-enjoyment purposes, such as for TDM, if an enjoyment purpose coexists with this non-enjoyment purpose, Article 30-4 of the Act does not apply.”¹¹ As a concrete example, it refers to “cases among incremental learning conducted on an existing pre-trained model (including the collection and processing of training data for that purpose), in which, intentionally and for the purpose of causing the model to output all or part of the original expression contained in the works included in the training data, a person reproduces works” (underlined by the author). It also notes that, even where such an intentionality does not exist, there are situations in which the use of “a small amount of training data to conduct incremental learning” may nevertheless fall under the coexistence type.¹² If this is followed, Article 30-4 of the Japanese Copyright Act shall not apply if the purpose of outputting the original expression of the source work used for AI training exists at the training stage.

However, with respect to this approach, there is also a view that, on the grounds that “information analysis” (TDM) is cited as an example of non-enjoyment use under Article 30-4(ii) of the Japanese Copyright Act, any use for the purpose of TDM should always be considered a non-enjoyment use.¹³

Learning,’ they permit only the training phase and do not in any way allow copyright infringement in the output (generation) phase. See Tatsuhiro Ueno, ‘「生成」と「学習」区別し対応を’ [Distinguish Between ‘Generation’ and ‘Training’ and Respond Accordingly], *Nihon Keizai Shimbun*, Feb. 26, 2024, at 16 (in Japanese); see also Ueno & Okumura (eds), *supra* note (1), at 71 et seq. [Ueno].

¹⁰ Regarding Article 30-4 of the Japanese Copyright Act, see Tatsuhiro Ueno, ‘The Flexible Copyright Exception for ‘Non-Enjoyment’ Purposes: Recent Amendment in Japan and its Implication’ (2021) 70(2) *GRUR International* 145-152.

¹¹ “General Understanding,” at 19 et seq.

¹² “General Understanding,” at 20.

¹³ See Yasuyuki Echi in Ueno & Okumura (eds), *supra* note (1), at 15 et seq.; the statement of Makiko Takabe at Copyright Subdivision of the Cultural Council, 69th Meeting (Mar. 19, 2024) (in Japanese)

However, in the Japanese Copyright Act after the 2018 amendment, provisions concerning “information analysis” (TDM) exist not only in Article 30-4(ii), but also in Article 47-5(1)(ii). Under said article, minor output of an original work used for AI training is permitted under certain conditions (paragraph 1 of the same article), and preparatory actions for that purpose are also allowed under certain conditions (paragraph 2 of the same article). From this, it is understandable that the “General Understanding” takes the position that, in cases where an enjoyment purpose coexists, namely, where the output of the original work used for AI training is contemplated, only said article applies, and Article 30-4 of the Japanese Copyright Act does not.

However, how to determine whether an enjoyment purpose coexists remains an issue. The “General Understanding” points out that, “even if an AI generates materials whose original expressions overlap with those of the works which the AI learned from at the generation or utilization stage, such a fact alone ordinarily does not allow one to infer the existence of an enjoyment purpose at the development or training stage.” At the same time, it states that “circumstances in which generated materials whose original expressions overlap with the works that have been learned from occur remarkably frequently at the generation or utilization stage may serve as one factor in inferring the existence of an enjoyment purpose at the development or training stage” (underlined by the author).¹⁴ Therefore, the mere fact that content shared between the original work used for AI training and the original expression occurs frequently as a result does not, by itself, establish the coexistence of an enjoyment purpose.¹⁵

Based on this, even if Article 30-4 of the Japanese Copyright Act does not apply in cases where an enjoyment purpose coexists, Article 47-5 of the same Act may still apply.¹⁶ Article 47-5 stipulates that “persons who engage in activities such as “using a computer to ... search for any information concerning the identification or location of information being searched for and making the results of that search available” (paragraph 1, item (i) of the same article) or “undertaking computerized data analysis and furnishing the results of that analysis” (item (ii) of the same paragraph), namely, those undertaking an action that “contributes to facilitating the exploitation of a work by creating new knowledge or information through computerized data processing, “may, with respect to publicly transmitted works, in any way and to the extent considered to be necessary in light of the purpose of the action, when exploiting it incidental to the undertaking of that action, undertake a minor exploitation” (the main text of the introductory clause of same article).

On this point, the “General Understanding” states that, with respect to “retrieval-augmented

(<https://www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/bunkakai/69/index.html>). See also discussion in Ueno & Okumura (eds), supra note (1), at 207 et seq.

¹⁴ “General Understanding,” at 21.

¹⁵ Note that in the draft of “General Understanding” (Dec. 20, 2023 edition), the term “remarkably” was absent. Consequently, this author pointed out that “business operators may be forced to consider that Article 30-4 cannot be applied unless they take measures to prevent at least ‘frequent occurrences’” (Ueno, Opinion, Jan. 2, 2024). In response to this, the term “remarkably” was inserted.

¹⁶ See “General Understanding,” at 10.

generation (RAG) and other methods in which generative AI searches target data that include works and produces an answer by summarizing the search results, etc.” (hereinafter referred to as “RAG, etc.”), “even in cases where Article 30-4 of the Act does not apply, it is conceivable that the use of existing works in the course of generating answers through RAG, etc., may fall under Article 47-5(1)(i) or (ii) of the Act”.¹⁷

Indeed, in the previous discussions as well, it has been explained that, in the case of plagiarism-detection services, the act of “digitizing a large number of articles, books, and other materials, analyzing the textual correspondence between such materials and the article to be examined, and, incidentally to providing information on whether there has been any copying from other works or on the degree of plagiarism, displaying portions of the original text of the works corresponding to the plagiarized passages,”¹⁸ or, in the case of medical-treatment recommendation services, the act of “analyzing various information on past case reports, treatment methods, drug efficacy, and the like, based on the patient’s condition, to determine the optimal treatment method and providing the results, and, together with the results, providing portions of the information that serve as the basis for determining the optimal treatment method,”¹⁹ falls under Article 47-5(1)(ii) of the Japanese Copyright Act. In light of this, it follows that the output of works accompanying so-called RAG may also be subject to the application of the same article, provided that the requirements of ‘incidental’ nature and ‘minor’ nature are satisfied. As various AI services are provided and developed in the future, the scope of application of said article will be a major issue in Japan.

(2) Proviso

The proviso of Article 30-4 of the Japanese Copyright Act (hereinafter referred to as the “proviso”) provides that “provided, however, that this does not apply if the action would unreasonably prejudice the interests of the copyright owner in light of the nature or purpose of the work or the circumstances of its exploitation.” Therefore, even a case falling under TDM will be excluded from the scope of said article if it falls under the proviso. In this regard, there was a lively debate on the following points, in particular, in the meetings of the Subcommittee.

(i) Targeted training

First, there is the question of whether a situation falls under the proviso’s cases where “the action would unreasonably prejudice the interests of the copyright owner” when a generative

¹⁷ “General Understanding,” at 21 et seq.

¹⁸ Moriyuki Kato, 著作権法逐条講義〔第7版〕 [Copyright Act Article-by-Article Lecture (7th rev. ed.)], (CRIC 2021) 419 (in Japanese).

¹⁹ Agency for Cultural Affairs, Copyright Division, ‘デジタル化・ネットワーク化の進展に対応した柔軟な権利制限規定に関する基本的な考え方（著作権法第30条の4、第47条の4及び第47条の5関係）’ [Basic Approach to Provisions for Flexible Limitations on Rights in Response to Advances in Digitization and Networking (related to Copyright Act arts. 30-4, 47-4 & 47-5)] (Oct. 24, 2019) 31 (in Japanese), available at https://www.bunka.go.jp/seisaku/chosakuken/hokaisei/h30_hokaisei/pdf/r1406693_17.pdf.

AI system trained extensively only on works by a specific creator (e.g., Studio Ghibli films, Akira Toriyama's manga, Beatles music, or Van Gogh paintings) produces a large volume of content that shares the same style, abstract manner of expression, or worldview as those works, thereby causing disadvantage to that creator.

Of course, while copyright protects concrete “expressions,” abstract “ideas” are not subject to copyright protection. Given this, there is no dispute that if the original expression of the source work used for AI training is not output, and only content that is merely common at the idea level is output, this does not constitute a copyright infringement of the source work used for AI training. However, some commentators argue that, at the stage of developing an AI system specifically targeted at a particular creator, conducting training that is specialized to that creator's works could potentially impair future potential sales channels for those works, and may fall under the proviso.²⁰

In this regard, the “General Understanding” presents the conventional view that, although “one can imagine situations in which a large volume of outputs that share only ideas or other elements that are not protected interests under copyright law result in AI-generated materials substituting for the demand for particular creators or works, such situations would not fall under cases where ‘the action would unreasonably prejudice the interests of the copyright owner’ under copyright law, in case the relevant generated materials do not share the original expression of the source work used for AI training.” Immediately thereafter, however, it notes that “from the perspective that the ‘interests of the copyright owner’ prescribed in this proviso are not necessarily identical to, and may be examined separately from, the damages arising from copyright infringement, a certain number of opinions held that, where demand for a particular creator or particular works is displaced by AI-generated materials, such situations could fall under cases where ‘the action would unreasonably prejudice the interests of the copyright owner’”.²¹

The wording of this section was the subject of intense debate within the meetings of the Subcommittee. This section is the only section where, as a result of repeated revisions, while the prevailing view was presented as the conventional view, the opinions of several members who took a more cautious position, in contrast to the prevailing view, were recorded in the main text rather than in a footnote, and were noted as “a certain number”.²²

However, even if a generative AI system exhaustively trained on the works of a particular creator outputs, in large quantities, content that shares the same style, abstract manner of expression, or worldview as those works, thereby resulting in economic disadvantage to that

²⁰ See Yasuyuki Echi, ‘AI 生成物・機械学習と著作権法’ [AI-Generated Materials, Machine Learning, and Copyright Law], (2020) 73-8 Patent 142 (in Japanese); Echi in Ueno & Okumura (eds), *supra* note (1), at 29 et seq. (in Japanese); Toshiya Kaneko, ‘生成 AI とせいせいしない著作権法’ [Generative AI and the Frustrating Copyright Law], (2024) 828 Hōgaku Seminar 58, etc. (in Japanese).

²¹ “General Understanding,” at 23.

²² For details, see minutes of the 6th Subcommittee meeting (Jan. 15, 2024) (in Japanese) (https://www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/hoseido/r05_06/) and the 7th Subcommittee meeting (Feb. 29, 2024) (in Japanese) (https://www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/hoseido/r05_07/).

creator, the cautious position remains open to question.²³ This is because, first, it is difficult to regard the action in question as harming an interest protected by copyright, and therefore it can scarcely be said to “unreasonably” prejudice “the interests of the copyright holder” as provided in the proviso. Second, although there is no dispute that reverse engineering of computer programs is a concrete example to which the introductory clause of Article 30-4 applies, one might wonder whether the proviso of the introductory clause of the same article would nonetheless be triggered in cases where the development of a program — incorporating common elements found in algorithms that themselves are not subject to copyright protection — results in economic disadvantage to the copyright holder of the original program used for training. Third, there is concern that such an interpretation could unduly chill the development of specialized AI models for which large-scale generation is envisioned.

(ii) Database works created for TDM activities

There was also intense debate regarding database works that have been organized in a form suitable for TDM.

This is because, under Article 47-7 before the 2018 amendment, the predecessor to Article 30-4(ii) of the Japanese Copyright Act, it was stipulated in the proviso that “this does not apply with regard to database works compiled for use by persons who carry out data analyses,” and thus, prior to that amendment, only database works created for TDM activities were subject to the proviso. Therefore, as an interpretation of the current Act as well, it has been explained that “when database works organized in a form that allows a large volume of information to be readily used for TDM are being sold,” they fall under the proviso of the introductory clause of Article 30-4 of the current Japanese Copyright Act.²⁴

However, because the proviso of Article 30-4 of the Japanese Copyright Act only provides that “this does not apply if the action would unreasonably prejudice the interests of the copyright owner,” there could also be a view that there is no limitation to database works created for TDM activities, and other cases could more broadly fall under the proviso.

However, considering that, in the legislative process of the 2018 amendment, non-enjoyment uses subject to Article 30-4 of the Japanese Copyright Act were positioned as “acts that do not normally harm the interests of the rights holder” (the so-called “first layer”),²⁵ it follows that cases where “the action would unreasonably prejudice the interests of the copyright owner” (underlined by the author) are exceptional. Moreover, the amendment expanded the provisions of limitations and exceptions on rights concerning TDM and did not intend to narrow the limitations and exceptions, as evidenced by the Diet’s additional resolution stating that “acts assumed to be subject to the limitations and exceptions on rights shall continue to be subject

²³ For details, see minutes of the 7th Subcommittee meeting (Feb. 29, 2024) [Statement of Ueno].

²⁴ Agency for Cultural Affairs, Copyright Division, *supra* note (19), at 9.

²⁵ See ‘文化審議会著作権分科会報告書’ [Report of Copyright Subdivision of the Cultural Council] (Apr. 2017) 41 et seq (in Japanese), available at https://www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/pdf/h2904_shingi_hokokusho.pdf.

to such limitations and exceptions”.²⁶ Also, considering that acts lawful prior to the amendment should be interpreted as not becoming unlawful after the amendment, among other factors, regarding the acts newly subject to rights restrictions under the amendment (e.g., acts other than “recording on a recording medium or adaptation” [e.g., transfer, public transmission], information analysis not performed by computer), even if there are examples falling under the proviso beyond “database works compiled for use by persons who carry out data analyses,” it seems reasonable to interpret that, regarding the acts that were lawful under the main text of the former Article 47-7, only “database works compiled for use by persons who carry out data analyses” fundamentally fall under the introductory clause of current Article 30-4.²⁷

In this regard, the “General Understanding” does not take a clear position on the above controversy,²⁸ but instead discusses concrete examples of cases in which “database works organized in a form that allows a large volume of information to be readily used for information analysis are being sold,” a situation as to which there is no dispute that the proviso applies.²⁹

First, the “General Understanding” states that “where, in addition to articles and similar content made available for users to view on a website, an API is offered for a fee that allows users to obtain data organized from a database work in a form readily usable for TDM, the act of reproducing, for the purposes of TDM and without paying for use of the API, a certain set of information from the articles and similar content posted on the website for viewing — where such set contains original expression of the database work — may fall under the proviso and therefore may not be subject to the copyright exception under this article”.³⁰ This can be understood as indicating that, where one collects individual article works without using the paid API for the database work created for analytical use and thereby effectively produces a reproduction of that database work, “there may be cases” in which such conduct falls under the proviso in relation to the database work created for TDM activities.

Second, the “General Understanding” states that “if technical measures have been implemented to prevent reproduction of works for AI training, and if, based on the fact that such measures are in place as well as past performance (such as the creation of works in a database organized in a form usable for TDM, or performance related to licensing transactions, etc.), it can be inferred that works in a database — including data on the relevant website — organized

²⁶ See House of Councilors, Committee on Education, Culture, Sports, Science and Technology (May 17, 2018) (in Japanese), available at https://www.sangiin.go.jp/japanese/qianjoho/ketsuqi/196/f068_051701.pdf.

²⁷ See Tatsuhiro Ueno, ‘平成30年著作権法改正について’ [On the 2018 Copyright Act Amendment], in Ryu Takabayashi, Ryōichi Mimura & Tatsuhiro Ueno (eds), *年報知的財産法 2018-2019 [Yearbook of Intellectual Property Law 2018–2019]* (Nihon Hyōronsha 2018) 4 et seq (in Japanese). For a contrary view, see Yasuyuki Echi, ‘AI生成物・機械学習と著作権法’ [AI-Generated Materials, Machine Learning, and Copyright Law], (2020) 73-8 Patent 131 (in Japanese).

²⁸ See also “General Understanding,” at 24 n.27.

²⁹ In other words, page 24 et seq. of the “General Understanding” present the sections titled “(c) Examples of works in a database organized in a form that can be utilized for information analysis” and “(d) Specific examples of the above (c) that may fall under this proviso (including the approach to cases, etc., where technical measures have been implemented to prevent reproduction for training purposes).”

³⁰ “General Understanding,” at 25.

in a form usable for TDM are planned to be sold in the future, then circumventing these measures and collecting a large volume of data from the website via a crawler in order to reproduce the database works for AI training may be considered, in relation to the database works, as falling under this proviso as an act that impedes potential future markets for the database works, and therefore would not be subject to the limitations on rights under Article 30-4 of the Act”.³¹ This can be understood as indicating that, where the existence of technical measures to prevent AI training or similar circumstances allow one to infer that a database work for analytical use is intended to be sold in the future, and where bypassing such technical measures and massively collecting individual article works effectively results in producing a reproduction of that database work for analytical use, such conduct may fall under the proviso in relation to that database work for analytical use and may constitute copyright infringement of that work.

Indeed, if reproducing a database work created for TDM activities for the purpose of TDM is understood to fall under the proviso, then in cases where the collection of individual article works results in what can be regarded as a reproduction of that database work created for TDM activities, such conduct would likewise fall under the proviso, and Article 30-4 of the Japanese Copyright Act would not apply. And if there is a sufficient likelihood that such conduct will result in a reproduction of the database work for TDM, then it may be considered that the act of collecting the individual article works already presents a risk of infringing the copyright in that database work for TDM. On that basis, it would be possible to bring a preventive claim under Article 112(1) of the Japanese Copyright Act, seeking to have such collection discontinued on the basis of the copyright in the database work for TDM.

However, this portion of the “General Understanding” presupposes an extremely complex scenario and was the subject of very intense debate within the Subcommittee,³² and, as a result of repeated revisions to its wording, it must be said that the final description has become quite difficult to understand. Accordingly, the following points — particularly those that are prone to misunderstanding — should be noted.

First, as noted above, the “General Understanding” discusses only cases where there is no disagreement that the provisionary clause applies, specifically “when database works organized in a form that allows a large volume of information to be readily used for information analysis are being sold.” If this is the case, for an article database to fall under this category, it must fall under database works under Article 12-2 of the Japanese Copyright Act, meaning it must “exhibit creativity in the selection or systematic arrangement of the information thereof,” and it must also fall under database works “organized in a form that allows a large volume of information to be readily used for information analysis.” Accordingly, for an article database for which the applicability of the proviso is at issue, the question turns on whether the article database falls under database works created for TDM activities in that sense.³³

³¹ “General Understanding,” at 26 et seq.

³² For details, see the minutes of the 5th Subcommittee meeting (Dec. 20, 2023) (https://www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/hoseido/r05_05/) and subsequent meetings.

³³ It should be noted that footnote 29 on page 25 of the “General Understanding” introduces examples of “licenses and APIs that are being provided for text and data mining,” but emphasizes that whether they fall

Second, for it to be said that the mass collection of individual article work leads to the reproduction of database work created for TDM activities, “similarity” and “dependence” in relation to the database work created for TDM are required. Moreover, for a similarity to be acknowledged here, the collection of individual article works must reach a level comparable to the originality involved in the “selection” of information in the database work prepared for TDM — that is, the originality in choosing themes for articles from among countless events in the world. Therefore, for collections whose eligibility under the proviso is at issue, the question becomes whether similarity in such a meaning can be acknowledged.

Third, even if the large-scale collection of individual article works leads to a reproduction of a database work created for TDM activities and could fall under the proviso in relation to that database work, this only means that it may constitute copyright infringement of the relevant database work.³⁴ On this point, as a result of repeated revisions discussed until the final stage, the “General Understanding” added the phrase “in relation to the relevant database work,” making clear that the proviso applies only to database works created for TDM activities and not to individual article works (that is, even if the proviso applies, it does not constitute copyright infringement of the individual article works).³⁵ It also stated that “even if an act is evaluated as hindering the potential future market for the relevant database work, it cannot be said to hinder the potential future market of the individual works contained therein, and thus does not fall under this proviso in relation to said individual works”.³⁶ Therefore, it should be noted that even if the mass collection of individual article works falls under the proviso and constitutes copyright infringement of the database work for analysis, it does not constitute copyright infringement of said article works.³⁷

under “database works organized in a form that allows a large volume of information to be readily used for information analysis” “requires consideration on a case-by-case basis” (this point was also revised repeatedly from the initial draft [Dec. 20, 2023]). At the very least, simply providing a standard article database for licensing for TDM is not considered to fall under this category (see minutes of the 5th Subcommittee meeting [Dec. 20, 2023] [Statement of Ueno]; the “opinion” stated in the second sentence of footnote 29 on page 25 of the “General Understanding”).

³⁴ See the statements of Yukihiro Miwa, Senior Specialist for Copyrights, Copyright Division, and the author at the Subcommittee’s 6th meeting (Jan. 15, 2024).

³⁵ Regarding the passage on page 27 of the “General Understanding” stating that “in relation to the relevant database work, it may fall under this proviso and thus not be subject to the limitations and exceptions on rights under Article 30-4 of the Act,” in Material 2-1 of the 7th meeting (Feb. 29, 2024) being the final Subcommittee meeting, titled “Regarding the General Understanding on AI and Copyright (Draft) as of Feb. 29, 2024 (Integrated Version),” the phrase “in relation to the relevant database work,” was absent (p.26 et seq.). It was ultimately added following repeated comments by the author (see minutes of the 7th Subcommittee meeting [Statement of Ueno]).

³⁶ “General Understanding,” at 27, n. 33. It is important to note that here it is described as a “pointing out” rather than an “opinion.”

³⁷ It should be noted that page 41 of the ‘AI と著作権に関するチェックリスト&ガイダンス’ [Checklist & Guidance on AI and Copyright] by the Agency for Cultural Affairs, Copyright Division (July 31, 2024) (https://www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/seisaku/r06_02/pdf/94089701_05.pdf) (in

III Copyright Issues at the Output Stage

1 Identification of Issues

Second, there is the question of whether content output by generative AI constitutes an infringement of copyright, etc.

When generative AI outputs content (e.g., images, audio, text), it does not infringe the copyright, etc., of others if said content is entirely new. However, when the content output by generative AI resembles copyrighted work, etc. (especially copyrighted works, etc., that said generative AI has previously learned from), the question of whether this constitutes an infringement of copyright, etc., becomes an issue.

In general, both similarity and dependence are required for an action to constitute copyright infringement.

Among these, similarity is judged based on whether the “essential expressive features” of another person’s works can be directly perceived, and for similarity to be acknowledged, it is necessary not only for the facts and ideas to be common, but also for the “original expression” protected by copyright to be common.³⁸ Therefore, similarity is not affirmed merely because content output by AI shares facts or ideas (including artistic style or manner) with another work; however, if it shares the original expression of the source work used for AI training, similarity is affirmed.

On the other hand, dependence means that a work was created by depending on another person’s work.³⁹ Consequently, a person’s inadvertent creation of the same work without depending on another person’s work does not constitute copyright infringement. However, in the case of humans, even if they have previously seen or heard another person’s work, dependence may be denied if it has been completely forgotten. In the case of AI, however, the question arises as to whether dependence should always be presumed for the vast number of works it has previously learned from.

Japanese), in explaining the “General Understanding,” describes “AI learning databases” and states: “As a rights holder, when publishing your works on the Internet, you can legally prevent unauthorized AI training by ensuring that the collection of your works, including the relevant work, is sold (licensed) as an AI learning database (dataset) for AI training or other TDM purposes.” However, if one follows the “General Understanding,” even if the rights holder takes such measures, they cannot prohibit such TDM based on the copyright of “their own works” referred to here, which raises concerns about potential misunderstandings.

³⁸ See Supreme Court, First Petty Bench, June 28, 2001, 55-4 Minshū 837 [Esashi Oiwake Case]. For details, see Tatsuhiro Ueno & Tetsuo Maeda, 〈ケース研究〉著作物の類似性判断——ビジュアルアート編 [Case Studies on the Determination of Similarity of Works: Visual Arts] (Keisō Shobō 2021) (in Japanese).

³⁹ For details, see Tatsuhiro Ueno, ‘著作権侵害訴訟における依拠性に係る要件事実’ [Requisite Facts Concerning Dependence in Copyright Infringement Litigation], in Shigeo Ito (ed), 知的財産法の要件事実 [Requisite Facts in Intellectual Property Law] (Nihon Hyōronsha 2016) 131 (in Japanese).

2 Previous Discussions

This issue was also discussed by the “New Information Property Examination Committee,” established in 2016 within the Intellectual Property Strategy Headquarters,⁴⁰ and the committee’s report states that “It has been questioned whether dependence can be affirmed merely because the original work is included in the training data, and even if dependence is affirmed, whether it is appropriate to acknowledge dependence in cases where the creator of the pre-trained model and the person who outputs the AI-generated materials are different. On this point, it is pointed out that if works included in the training data, either in whole or in part, are output by the pre-trained model, and dependence were uniformly denied, then, it would effectively allow copyright infringement to be denied simply by using AI. Consequently, it is conceivable that AI could be used for the purpose of committing copyright infringement, or that one might falsely claim to have used AI to evade liability for infringement even when AI was not actually exploited. Against the backdrop of these points, it has been pointed out that dependence should be acknowledged when a work is retained as is, as data in a pre-trained model in the form of creative expression. It has also been pointed out that even if the work is not retained as is, dependence may be acknowledged as long as there was access to the original work, such as its inclusion in the training data, and that the determination of infringement could then be based solely on similarity. On the other hand, it has been pointed out that when a work is abstracted or fragmented as parameters rather than as original expression, only the underlying ideas are being used, and dependence should not be acknowledged. In addition, when considered in parallel with dependence in human creativity, if mere access to a work were sufficient to acknowledge dependence, the defense of independent creation under the Japanese Copyright Act would no longer function, potentially narrowing the space for freedom of expression.”⁴¹

Indeed, while it may seem unavoidable to acknowledge dependence on works that an AI has learned from, from the perspective that such a view could expose AI users to unforeseen legal risks, there is also a view that the possibility of denying dependence should be explored.⁴²

⁴⁰ In particular, see minutes of the 2nd (Dec. 5, 2016)

(https://www.kantei.go.jp/jp/singi/titeki2/tyousakai/kensho_hyoka_kikaku/2017/johozai/dai2/gijiroku.pdf) and 3rd (Dec. 19, 2016)

(https://www.kantei.go.jp/jp/singi/titeki2/tyousakai/kensho_hyoka_kikaku/2017/johozai/dai3/gijiroku.pdf)

meetings of the New Information Property Examination Committee.

⁴¹ ‘新たな情報財検討委員会報告書——データ・人工知能（AI）の利活用促進による産業競争力強化の基盤となる知財システムの構築に向けて——’ [Report of the New Information Property Examination Committee] (March 2017), 37 et seq (in Japanese). Available at

https://www.kantei.go.jp/jp/singi/titeki2/tyousakai/kensho_hyoka_kikaku/2017/johozai/houkokusho.pdf.

⁴² Okumura in Ueno & Okumura (eds), supra note (1), at 120 states that “in our view, one should distinguish and discuss based on whether what is retained in the pre-trained model as a result of learning constitutes an expression or an idea; in the former case, dependence at the generation stage is acknowledged, whereas in the latter case, dependence at the generation stage is not acknowledged.”

3 “General Understanding”

On this point, the “General Understanding” presents the conventional view by stating that, “Even if an AI user was not aware of existing works (their expressive content), if the AI learned from those works during its development and training stage, it can be objectively acknowledged that there was access to those works. Therefore, when the AI is used and generated materials similar to those works are generated, dependence is generally presumed, and it could constitute copyright infringement by the AI user.” At the same time, it states that “However, there may also be cases where it can be said that, with respect to the relevant generative AI, it is technically ensured that the original expressions of works used for training during its development and training stages will not be generated at the generation or utilization stage. In light of circumstances such as the fact that such a state is technically ensured, if it can be legally assessed that the generative AI is not in a state to output the original expressions of works used for training during the generation or utilization stage, AI users may assert the underlying facts to support that assessment, and there may still be cases in which no dependence is deemed to exist, even if the generative AI learned from existing works during its development and training stages.”⁴³

In this way, the “General Understanding” presents, as a conventional view, the position that dependence is, as a rule, affirmed with respect to works used for training the AI, while at the same time suggesting that dependence may be denied in certain circumstances — a point that deserves particular attention. However, the question of the specific circumstances under which such dependence can be denied will require further examination.

IV Concluding Remarks

Amid growing attention and concern from various quarters regarding AI and copyright, the “General Understanding” sets out the conclusions of discussions reached on the relevant issues and can be said to be meaningful for a wide range of stakeholders.

However, the “General Understanding,” as stated on its cover, “presents certain views of this Subcommittee at the time of its publication,” and, as it notes, “going forward, since the accumulation of concrete cases, the development of AI and related technologies, and the progress of examinations in other countries are expected to continue, we will continue to strive to collect and grasp relevant information and, as necessary, review this General Understanding and conduct other necessary examinations.” Thus, it merely presents the Subcommittee’s views at the present time on a provisional basis.⁴⁴

In the future, further technological and social developments, as well as the evolution of

⁴³ “General Understanding,” at 34.

⁴⁴ Given this, the Secretariat of the Subcommittee had suggested even at the final stage that the title should retain the designation “(素案)” [(draft)] (see minutes of the 7th Subcommittee meeting [Feb. 29, 2024], [Statement of Keiko Momii, Director, Copyright Division, Agency for Cultural Affairs]).

international discussions, are expected, so it will be necessary to continue deepening the discussion on this issue, including a re-examination of the current “General Understanding.”



The WIPO Japan Office supports this article under a Funds-In-Trust Japan Industrial Property Global (FIT/Japan IP Global) project.