

Author:

[E-mail✉ Shinnosuke Fukuoka](#)

[E-mail✉ Tomonobu Murata](#)

[E-mail✉ Atsuki Mizuguchi](#)

1. Introduction

Recently, generative AI (“GenAI”)¹ has become a major topic of discussion around the world, and Japan is no exception. GenAI can increase human productivity greatly, but there are legal issues involved in its development and use that individuals, and businesses in particular, should understand. As such issues are vast,² given the wide ranging impacts of GenAI, this series of newsletters will focus on one area of Japanese law at a time, beginning with copyright-related issues.

2. Four Stages of Generative AI Use

Copyright infringement issues arise when using someone else’s work as “training data³” or “prompts,⁴” or when AI-generated material resembles someone else’s work. Appropriate exploration of such copyright issues requires separate examination of GenAI’s four stages of use:

(1) Model Training

- (1.a) creating and training a machine learning model with a large training data set; and
- (1.b) further training the model with data in a specific field (fine tuning).

(2) Model Using

- (2.a) user prompts (e.g., chat input or images); and
- (2.b) AI output based on prompts.

It can be confusing to discuss these stages without distinguishing between them; thus, it is appropriate to discuss them each in turn.

¹ GenAI is based on artificial neural network technology, comprising numerous parameters (e.g., internal values). Two types of GenAI have been attracting attention in recent years: “large language” models (such as those used in services like ChatGPT, Bing AI, Bard, etc.) and “image generation” models (such as those used in services like Midjourney, Stable Diffusion, Firefly, etc.), but the legal concerns are the same in principle. GenAI which imitates a specific person’s voice also bring into consideration legal issues regarding the neighboring rights of performers (e.g., voice actors).

² The risks include: (i) copyright infringement, (ii) use of erroneous information, (iii) leakage of confidential information, (iv) inappropriate use of personal information, and (v) misuse. Problems also arise when GenAI generates incorrect information that humans use to make judgments or take actions. If confidential or personal information is used in training data or prompts, leakage of confidential information or inappropriate use of personal information also becomes a problem. Misuse of GenAI to create fake news or viruses also is an issue.

³ Training data – data fed to an AI model (or machine learning algorithm) in order to allow it to make decisions based on the information therein.

⁴ Prompts – questions, coding, information, or text of any form used to interface with an AI or machine learning system.

Parameters in the model are tuned in the training process of Stage (1.a) so that the model can produce output with characteristics or regularities similar to the training data. Stage (1.b) also is part of the training process, but it uses a more focused data set in order to make the AI model more specialized in the specific domain, such as medical, law and advertisement. Since stage (1.a), in which models are trained with a very large data set, requires advanced AI technical capabilities and significant financial resources, typical companies should focus on the issues in stages (1.b), (2.a), and (2.b). Stage (2.a) and (2.b) are usually conducted by users of AI services, so such users and services should focus their attention on the legal issues surrounding these stages.

3. Whether Use Constitutes Copyright Infringement

(1) Model Training Stage

In the training phase (1), the legal issue is the use of another person's copyrighted work as training data, i.e., whether or not the copy, transfer, etc., of the work infringes the copyright thereof.

Some countries have instituted comprehensive exceptions that limit copyrights, such as Fair Use (e.g., 17 U.S.C. Article 107). The issue then falls to whether such provisions are applicable to the use in question. However, the Copyright Act of Japan (Act No. 48 of 1970; unofficial English translation is available [here](#)) does not provide a comprehensive exception, but rather specific provisions in Articles 20 to 50. Amendment to the Copyright Act of Japan in 2018 added an exception applicable to AI training in Article 30-4, making it one of the more "relaxed" copyright acts in the world under which business would likely be allowed to use copyrighted training data for AI development unless exceptional requirements are met by those claiming infringement. Accordingly, it is advisable for service providers of GenAI to familiarize themselves with this article's concepts and applicability.

Article 30-4 of the Copyright Act (below) permits the use of a copyrighted work without the permission of the copyright holder to the extent deemed necessary, provided that the purpose is not for oneself or others to "enjoy" the thoughts and feelings expressed in the work. Item 2 of the same provision lists "information analysis" as an example case where the purpose is not to have humans "enjoy" the work.

(Use not intended for the enjoyment of ideas or emotions expressed in a work)

Article 30-4: *A work may be exploited, regardless of the method used, in the following cases and other cases where the purpose is not to enjoy the ideas or sentiments expressed in the work for oneself or to have others enjoy the work, to the extent deemed necessary for such purpose: provided, however, this shall not apply where, in light of the type and intended use of the work and the manner of such use, such use would unreasonably prejudice the interests of the copyright holder:*

- (i) for the purpose of testing for the development or practical application of technology pertaining to sound recording, video recording or other exploitation of a work;*
- (ii) for the purpose of information analysis (meaning the comparison, classification or other analysis of information pertaining to language, sound, images or other elements constituting said information extracted from a large number of works or other large quantities of information; the same shall apply in Article 47-5, paragraph (1), item (ii));*
- (iii) in addition to the cases listed in the preceding two items, when the work is used in the course of information processing by computer or for other use (in the case of a program work, excluding computer execution of the work) without any human perceptual recognition of the expression of the work.*

* Emphasis added.

Using another person's work as training data for AI usually falls within this "information analysis" category as it does not aim to create enjoyment of the ideas or sentiments expressed in the work, which means that the work may be used without the permission of the copyright holder, as long as the use falls within the necessary scope.

However, if the purpose of the training involves enjoyment of the ideas or sentiments expressed in the work, for example, making a new creation from which we can feel essential characteristics expressed in the work, Article 30-4 of the Copyright Act is not applicable. For example, some fine tuning methods such as LoRA (Low-Rank Adaptation) enable models to output a specific character or art style with training that involves less data. In this case, careful legal analysis would be desirable to check whether the requirement above is satisfied.

Moreover, Article 30-4 of the Copyright Act states that permission of the copyright holder is required "when the interests of the copyright holder are unreasonably impaired" based on the type of work to be used, its intended use, and the manner of use. This provision reconciles the interests of the copyright owner and the user. Whether a use is "unreasonably prejudicial to the interests of the copyright holder" is determined on a case-by-case basis. Informatively, the Copyright Division of the Agency for Cultural Affairs' "[Basic Idea on Flexible Rights Restriction Provisions Responding to the Advancement of Digitalization and Networking](#)" states that whether a situation constitutes "unreasonable harm to the interests of a copyright holder" is evaluated from the perspective of (i) whether it conflicts with the market for the copyright holder's works, or (ii) whether it will hinder the potential market for the works in the future. For example, if a database that organizes a large amount of information so that it can be used easily for information analysis is available for sale, unauthorized use of the database as learning data constitutes "unreasonable harm to the copyright holder's interests."

However, since "unreasonable harm" is an abstract concept, it is unclear at this point what constitutes it.

In addition, when using data from a website as training data, if the website has Terms of Use that include phrases such as "data from this website may not be used for commercial purposes" or "data may not be used for AI training," or if data that is prohibited by contract from being used for other purposes is used as training data, it is not clear whether it is permissible to use the data in violation of the terms of use or contract. In such cases, there also is the issue of "copyright override," which is the question of whether it is acceptable to use the data in violation of the terms of use or contract.

(2) Model Using Stage

(a) User Prompts

In the prompt input phase, a problem arises when a user uses someone else's copyrighted work in a prompt. In addition to images, such as illustrations, and texts, such as articles, prompts created by others can be considered other people's copyrighted work.

The question of whether Article 30-4 of the Copyright Act applies to the use of another person's work as a prompt also is an issue, meaning the details and limitations discussed in the prior section could apply. However, the use of copyrighted works as a prompt does not generally constitute "information analysis" under

Article 30-4(2) of the Copyright Act, since it does not require the reading of a large amount of data. A more important issue may be whether or not it aims to create enjoyment for oneself or to have others enjoy the ideas or sentiments expressed in the copyrighted works of the prompt. For example, if a user puts an image or the name of their favorite character into a prompt with the aim of generating an image of their favorite character, the prompt may be deemed to be intended to be enjoyed by a human being. In addition, depending on the input content of the prompt, it may constitute "unreasonable harm to the interests of the copyright holder."

(b) AI Output

In the output phase, copyright infringement becomes an issue when the output is similar to someone else's work. Copyright infringement is established when the output (i) is similar to someone else's work ("similarity") **and** (ii) relies on someone else's work ("reliance").

Similarity

The way to judge similarity between a copyrighted work and an AI output is considered the same as that between a copyrighted work and a human product.

Reliance

On the other hand, with respect to the reliance between a copyrighted work and an AI output generated from a model using the work in the training data, there is some debate as to whether a copyrighted work can be said to have been "relied upon", since the work may be only a small portion of a vast amount of training data and users necessarily do not recognize what kind of training data was used for an AI. Broadly speaking, there are two opinions: one is that the data is not relied upon because it is only an idea in fragments, and the other is that the data is relied upon because it is included in the data for learning purposes.

However, if a work is used as training data for fine tuning with an intention to imitate its expression, there is a higher possibility that the AI output is deemed to rely on the work. More specifically, if a user uses another person's image or other copyrighted work as a prompt, it will be clear that reliance occurred. If a user enters the name of an artist or character as a prompt, e.g., if you enter "Please create a sentence in the style of Haruki Murakami" or "Please create a character in the style of Pokemon" as a prompt, although you are not directly using the artist's work or character images, reliance could be deemed to have occurred, because it is conceivable that this prompt may lead to the creation of a Haruki Murakami-like sentence or a Pokémon-like character based on a sentence by Haruki Murakami or a Pokémon image learned by the GenAI.

Other Requirements

If the copyright infringement is deemed to occur through satisfaction of requirements (i) and (ii) above, the copyright holder is entitled to demand an injunction against the user's use. However, since users do not necessarily know what kind of copyrighted work is used in large training data sets, a claim for damages requiring willful misconduct or negligence might not be admitted unless the user intentionally instructs the AI to generate outputs similar to the copyrighted work or uses such data after the generation while acknowledging that it is copyrighted work.

Based on the "private use" exception (Article 30 of the Copyright Act), if an individual only creates a prompt and views or downloads the output, this does not constitute copyright infringement. However, if the output is

published on the Internet or such creation or downloading is conducted by an employee for business operations of the employer, this private use exception is not available. Accordingly, it is advisable for businesses to establish internal rules on how to legally use GenAI services for business operations.

In addition, although the person who gave directions to the GenAI is basically considered a "user," an AI service provider might be responsible for infringement caused by the user under special circumstances where the service provider is deemed to engage or participate in the infringing behavior, etc.

4. Is An AI-generated Product Copyrighted?

While the Copyright, Designs and Patents Act 1988 of U.K. stipulates a "computer-generated" work, which means the work is generated by computer in circumstances such that there is no human author of the work, can be copyrighted under the act, the Copyright Act of Japan does not stipulate whether AI output can be copyrighted.

It is commonly accepted that copyright arises when there is human creative intent and creative contribution, since in Japan the main purpose of the copyright act is considered to protect creation. Since simple instructions given to an AI usually do not constitute creative intent and creative contribution, the AI-generated product cannot be copyrighted, and the human who gave the instructions will not have a copyright in it. In such a case, the copyright of the AI-generated product does not belong to anyone, and anyone is free to use it.

On the other hand, if a human being has a vision of an image or text that he/she wants to create using AI, and has taken the time and effort to realize that vision by devising prompts, or other efforts, copyright may be recognized on the basis of creative intent and creative contribution.

Therefore, if someone imitates an AI-generated image, whether a copyright can be claimed depends on how much human involvement occurred in the creation of the image, which will be determined on a case-by-case basis. It is expected that AI-based creativity will increase in the future. In order to claim copyright based on their own creativity, it will be important for creators who use AI to record the creation process and other relevant details so they can prove their creative intent and creative contribution.

5. Applicability of Copyright Act of Japan

As mentioned above, Article 30-4 of the Copyright Act enables service providers to conduct machine learning relatively free of legal issues. However, the concern of foreign readers may be whether Japanese laws are applicable to their use of AI.

In general, regardless of whether it is a claim for injunction or damages, the determination of whether copyright infringement has occurred is governed by the laws of the country where the alleged infringing exploitation took place.

The issue is the interpretation of the country home to the AI model exploitative act, especially when the location of the server providing the model differs from the location of the AI users (e.g., where foreign companies remotely provide users in Japan with GenAI services). There is no definitive consensus on this matter.

It is generally acknowledged that the location of the server is a relevant factor in determining the country of the exploitative act. Therefore, generally speaking, if a foreign service provider uses training data for AI development in a server located abroad, it would unlikely be able to rely on Article 30-4 of the Copyright Act even if the users are located in Japan, while service providers developing AI in Japan whose users are located abroad would likely be subject to Article 30-4 of the Copyright Act.

However, some opinions suggest that merely having a server in Japan does not automatically designate Japan as the country of the exploitative act. When considering the application of exception Article 30-4 in the Copyright Act of Japan for activities such as training, it is necessary to carefully examine the applicable law based on the specific circumstances and manner of utilization.

6. Conclusion

When using generative AI, it is important to be mindful of intellectual property rights, including copyright. Each country's copyright acts handle various issues related to copyrighted works in AI differently. However, it is noteworthy that, as mentioned above, the Copyright Act of Japan explicitly includes provisions that are accommodating towards AI training. Having said that, please note that there are various legal discussions and issues surrounding this topic as introduced above. If you contemplate using generative AI for your business in Japan, it would be advisable to seek legal opinions and assessment in advance.

In order to respond to the business needs of our clients, we publish newsletters on a variety of timely topics. Back numbers can be found [here](#). If you would like to subscribe to the N&A Newsletter, please fill out [the N&A Newsletter subscription form](#).

This newsletter is the product of its authors and does not reflect the views or opinion of Nishimura & Asahi. In addition, this newsletter is not intended to create an attorney-client relationship or to be legal advice and should not be considered to be a substitute for legal advice. Individual legal and factual circumstances should be taken into consideration in consultation with professional counsel prior to taking any action related to the subject matter of this newsletter.

Public Relations Section, Nishimura & Asahi [E-mail](#) 