



The use of copyrighted works in data sets for machine learning - Commentary on the 2018 reform of the Japanese Copyright Act Shinnosuke Fukuoka, Shu Numazawa

Reform of the Japanese Copyright Act in 2018 (the “amendment”) was largely undertaken in order to facilitate the progress of digitalization and data usage, improve access to information for persons with disabilities, and increase the utilization of archives. In this newsletter, we focus on the first of these goals and one of the mechanisms by which it is to be achieved: limitation of copyright protections in cases where the use “*is not for the purpose of enjoying expressed thoughts or emotions.*” After discussing this “loosening of regulation,” we describe how it will affect the development and use of AI in Japan.¹

1. What is the provision on the limitation of Copyright?

Generally, during the development and use of AI, the greater the access to data (i.e. fewer restrictions), the more accurate the resultant model. For example, when teaching an AI to identify an image of a “cat,” there are advantages to having as many cat sample images as possible: the AI can use the information to incorporate a wider number of shapes, colors, patterns, and so on into its definition of “cat.” As such, if an AI were allowed to utilize a large number of cat sample images from the internet, the accuracy of the AI’s determination process would increase. Unfortunately, it is likely that most such “cat” image data on internet websites is copyrighted, which (under the original version of the Japanese Copyright Act (“JCA”)) would prevent the use of the images for training an AI (i.e. machine learning).

Under Article 2(1), Item 15 of the JCA, saving copyrighted photos in data storage² constitutes “reproduction,” and it is, therefore, likely to infringe the “right of reproduction” (JCA Art. 21) of the copyright owner. However, in order to harmonize the economic

¹ The article numbers referred to in this newsletter are references to the Japan Copyright Act (the “JCA”).

² Generally considered a necessary step in the type of machine learning discussed in this document.

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.

interests of copyright owners with the potential benefits garnered by society's use of such information,³ the JCA will now (thanks to the reform) limit copyright protections in a manner that promotes machine learning, eliminating any associated claims of infringement. This change was necessary because the JCA has no comprehensive provisions on the limitation of copyright (such as "fair use" in the US system), so in order to use a copyrighted work (the "work") without the permission of the copyright owner, one must rely on the new "limitation" provisions of the amendment.

2. Background to this reform

Since AI (and the associated need for "big data") is continually evolving and becoming more common place, one of the purposes of this amendment is to allow the "flexible" use of copyrighted material in a manner that does not significantly affect the market for the work. According to the explanatory papers of the Agency for Cultural Affairs,⁴ there are three categories of such utilization: (A) that which would not normally harm the interests of the copyright owner, (B) that for which the disadvantage to the copyright owner is slight, and (C) that which may conflict with the copyright market but is expected to promote the use of the work for the realization of public policy. The amendment creates provisions on the limitation of copyright that conform to each category. Article 30-4, discussed here, falls under category A. Since category A is intended to be the most cautious when considering copyright owners rights, the language of Article 30-4 is somewhat ambiguous (i.e. reserved) with regard to the limitation of copyrights.⁵

3. Reformed Article

On January 1, 2019, the following amendment to JCA Article 30-4 was enacted (**emphasis** added by the authors):

Article 30-4 (Utilization not intended for the enjoyment of thoughts or emotions expressed in the work)

Where the intention of utilizing a work is not to enjoy the thoughts or emotion expressed in the work, or for another person's enjoyment or the enjoyment of others, or in the cases listed below [...] **the work can be used, provided [...] this shall not apply where it would unfairly harm the interests of the copyright owner** in light of the type and use of the work [...].

- 1 [...] testing for the development or practical application of technology relating to recording or other use of the work.
- 2 Information analysis (meaning to extract information on the language, sounds, shadows and other elements constituting the information from a large number of works and other large amounts of information and to perform comparison, classification and other analysis [...]).
- 3 [...], the use [...] of the work in information processing by [...] computer without the expression of the work being perceived by human beings [...].

³ Nobuhiro Nakayama "Copyright Law (2nd edition)" 281-282.

⁴ Copyright Center of the Ministry of Cultural Affairs of the Ministry of Education, Culture for Agriculture, April 2, 2018 "Legal Explanation Draft Outline Explanatory Material for Amendment of Copyright Law (Relating to AI Use Promotion)" https://www.kantei.go.jp/jp/singi/titeki2/tyousakai/kensho_hyoka_kikaku/2018/contents/dai4/siryou6.pdf

⁵ In addition, the limitation provision (Article 47-4), relating to the usage associated with copyrighted materials in electronic computers, is also classified as category A.

(1) When “*the purpose is not for the enjoyment of thoughts or emotions...*”

Firstly, in Article 30-4 (*amended*), it is stated that Items 1 to 3 (i.e. “*cases listed below*”) do not fall under the enjoyment of “*thoughts or emotions.*” This wording, as noted, is intentionally vague so as to increase the comprehensive nature of the provision. In light of this intention, it is generally understood that Items 1 to 3 are merely examples. According to the Cultural Agency explanatory papers, the wording “*where the intention of utilizing the work is not to enjoy the thoughts or emotions expressed in the work,*” excludes use of the work in a form that human beings “perceive”. Given this comprehensive nature of the provision, it is not clear whether Article 30-4 applies in specific cases. For clarification, we must wait for further discussion and the accumulation of judicial precedent.

(2) Item 1: Test use

One aim of Item 1 is to target the copying of artworks for purposes related to the development of cameras and printers.⁶ For example with regard to AI, Item 1 could be used for testing recognition of sensor output. Although Item 1 replaces the previous Article 30-4, some changes, such as those for unpublished works, have been made.

(3) Item 2: Use for information analysis

Item 2 inherited its impetus from the previous Article 47-7 which was originally intended to be the JCA’s primary connection to information analysis and the development of AI. The previous Article 47-7 was instead interpreted to mainly apply to “statistical analysis.” In the reform, Article 30-4 simply considers “analysis” to include “algebraic” and “geometric” functions, which may be used for analysis in AI development. Essentially, the language of Item 2 is drafted in a way that implies that gathering information for the purpose of analysis (in general) is permitted.

In addition, the previous Article 47-7 prescribed the use of storage for only recording mediums and adaptations. In Article 30-4 (*amended*) this restriction was removed so that information analysis concerning the work can be transferred to a third party or disclosed on a website (if the necessary requirements are satisfied). This allows engineers and researchers to share data for information analysis, and as a result, improves the usage of data for AI development.

(4) Item 3: Use without the expression of the work being perceived by human beings

Item 3 is a new provision that broadly covers the use of a work that does not involve the actual image being “perceived” by a human being. The scope of this addition to the law can be interpreted widely. Although the provision stipulates that the use must be without the perception of human beings, it is not obvious what level of “non-perception” is required (e.g. this could mean that while the AI analyzes an image the computer operator will not have the ability to see it, or something more or less stringent), so specific matters will need to be determined on a case by case basis.

(5) Proviso

The proviso of the amendment are understood as serving the same purpose as was prescribed in the previous Article 47-7. This

⁶ As discussed in the associated Diet deliberations.

proviso was drafted such that it would not unreasonably harm the interests of copyright owners and facilitate a general understanding that there has not been a substantial change from the previous Article 47-7. However, since the previous Article 47-7 only stipulates that “it does not apply in the case of a copyrighted database prepared for use by those who conduct information analysis,” it seems at first glance, that there is a change in the content of the proviso. None-the-less, it is reasonable to assume that such change is not substantive. It is believed that Article 30-4 does not change the status of any acts that were illegal under the previous Article. In any case, in interpreting the provisions of Article 30-4, if its application meets the requirements of the proviso of the previous Article 47-7, it can be said that the rights of the copyright owner are not unduly harmed.

4. Conclusion

As mentioned above, Article 30-4 was introduced to satisfy new needs for the use of copyrighted work by AI, but it is drafted in broad terms. It is important to make a determination as to whether the Article is applicable only after understanding the actual way in which the work will be used, while taking into consideration previous discussion and previous provision on the limitation of copyright.

In the previous Article 47-7, it was considered that there were no *onerous* restrictions on the use of data for machine learning in Japan. Due to the implementation of Article 30-4 by this amendment, the ability to use copyrighted works has been widened. As a result, it is expected that the use in Japan of copyrighted works (including those from foreign countries) for machine learning will increase. In such case, the use of cross-border works will be an intriguing issue, which we look forward to discussing in another newsletter.



Shinnosuke Fukuoka

Partner

E-mail: s_fukuoka@jurists.co.jp

Shin Fukuoka is a partner of N&A. In the area of Robotics/Artificial Intelligence, he mainly handles A.I., Big Data, and IoT. He has also contributed to many publications, including “Law and Contract of Data,” “Artificial Intelligence: Law and Issues,” and “Law and Strategy of the Internet of Things and Artificial Intelligence.”



Shu Numazawa

Attorney-at-Law

E-mail: s_numazawa@jurists.co.jp

Shu Numazawa was admitted as an attorney in 2015. He co-authored a book entitled “Artificial Intelligence: Law and Issues”, among others. Shu handles transactions and disputes relating to intellectual property rights and/or technology (including AI, automobile driving-related technology, etc.), data protection matters, and general corporate matters including venture support.