

Japanese Group's Provisory Response dated 3 June 2023 (Lacking Part 4 and 5)

Paris Congress

ALAI 2023

Artificial intelligence, copyright and related rights

June 22-23, 2023

To National Reporters:

The questionnaire uses the neutral term AI "production" to refer to content generated by an artificial intelligence system. As opposed to the term "work (of the mind)" which is the one that describes the classical object of copyright protection. This means that the content we are interested in is content produced by the artificial intelligence machine (or "system"), itself fed upstream by works of the mind, reproduced in a training data base. The margin of intervention of the final user is thus a priori very limited, but not always non-existent. The hypothesis concerned by this Congress is thus closer to what the ALAI once studied as "computer-generated creations" than to "computer-assisted creations" (see the 1989 Quebec City Congress).

In the mind of the editors of this questionnaire, an "artificial intelligence system" is defined as a computer system that allows, with a certain autonomy, automated decision making or predictions influencing real or virtual environments¹.

The questions raised are numerous because of the disruptive nature of the phenomenon, the multitude of issues and the theoretical, economic and social importance of the stakes.

Some of the questions will undoubtedly be accompanied by brief negative answers, which is already a useful answer for the General Reporters. Simply indicate these ("no", "none").

In other cases, the answers may be uncertain. In these cases, it is easiest to follow the classic pattern: "1) What do statutes and regulations say? 2) What does the caselaw say? 3) What does the national group think? To questions 1 and 2 above, the answer will often be "Nothing specific about AI but the relevant reference text/principle might be ...". Regarding 3), the national group is not obliged to have taken a position.

It is of this uncertainty and diversity that we will try to draw together, in June, a clear picture.

The team of the Scientific Committee (Alexandra Bensamoun, Jane Ginsburg, Silke von Lewinski, Pierre Sirinelli) is of course at your disposal to explain a question that might not seem, because of the particular context, immediately clear.

Thank you all and we look forward to seeing you in Paris.

Note: the questionnaires must be returned by the national groups no later than May 8, 2023. They will be sent to Pierre Sirinelli (pierre.sirinelli@univ-paris1.fr) and Sarah Dormont (sarah.dormont@u-pec.fr).

¹ This definition is comparable to the one retained by the European Union in the discussion on the AI Act (proposed regulation COM(2021) 206 final, March 2023 position), itself inspired by the 2019 OECD Recommendation on AI.

Artificial intelligence, copyright and related rights

The contours of the relationship

1. Understanding

1.1 - Has your national or regional law adopted a legal definition of AI?

No.

1.2 - Can you provide some examples of current uses of AI and its productions in the cultural sector of your country?

AI learned the singing style and appearance of a deceased, very famous singer and created a video of singing a new song.

In the field of music, an auto-composing service "FIMMIGRAM" exists (<https://fimmigrm.com/en/>). In the field of illustration or Manga, a creators' assistance service "mimic" exists (<https://illustmimic.com/en/about/>). But it accepts solely their users' own illustrations for the purpose of constituting training data.

1.3 - (Optional) What are the issues that have been exposed in your country on this subject: stakes, difficulties, orientations, proposals...?

Ethical issues of regenerating individuals by AI. Concerns by copyright holders such as illustrators and news publishers about the unauthorized training of their works for TDM.

1.4 - Are there any initiatives in your country or region aimed at regulating the use of AI in the cultural sectors?

The JDLA (<https://www.jdla.org/en/en-about/>), a non-profit organization composed by Japanese AI Startups aiming to promote Deep Learning technology, just has published May 1st its "Model Guideline for Using Generative AI".

This Model Guideline has a general purpose and is not specific in the cultural sectors. It is not yet translated into English.

It concerns only about ChatGPT and it consists mainly of 2 parts : prohibition of use for certain purposes (optional in each companies) and good practices for using. The latter part consists of 2 sub-parts: how and what to input in AI and how to use AI-generated output.

2. Understanding the upstream

2.1 - Are the AI system or its components likely to be protected by intellectual property rights (copyright and/or industrial property – patents, trade secrets . . .) ?

It is not clear but AI system and/or its components can be protected as work of computer program or database under the Japanese Copyright Act, in so far as it is regarded as 'creative expression' (Art. 2(i)(i)). A technical idea behind the AI system or its components can be protected by patent as an invention of product or method. Also a bigdata can be protected as 'protected data' under the Japanese Unfair Competition Prevention Act (Art. 2(1)(xi)-(xvi)), even if it is not regarded as work of database.

2.2 - Can rights under copyright be enforced against the use of protected contents by AI training?

Does the insertion of a pre-existing work into the computer system implicate rights under copyright?

If a 'machine reading' of works or protected subject matters is transient, it is excluded from the concept of reproduction (Art. 2(1)(xv)) and the right of reproduction (Art. 21) does not cover such machine reading thereof.

If so, in order to avoid a finding of infringement, are the copying or storage covered by an exception?

The Japanese Copyright Act has TDM exception that was first introduced in 2009 (Art.47-7) and was extended by the amendment in 2018, which came into effect in 2019 (Art.30-4(ii)), under which not only copying or storage of works but also even communication to the public of works can be allowed, in so far as it is conducted solely for TDM and necessary for it.

Article 30-4 (Exploitation without the Purpose of Enjoying the Thoughts or Sentiments Expressed in a Work)

It 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:

(ii) if it is done for use in data analysis (meaning the extraction, comparison, classification, or other 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; the same applies in Article 47-5, paragraph (1), item (ii))

**Underlined by National Reporters*

2.3 - In your country, are there any proposals to change the law and in which direction?

For example, by deeming that the incorporation of preexisting works into AI systems does not create an actionable "reproduction" of the works? Or by creating a new exception? Or by implementing a compulsory licensing system? Other solutions?

There is basically no discussion or proposal in Japan to change the Japanese TDM copyright exception at this moment. But there might be an opinion to add remuneration right for the TDM activities for commercial purpose.

2.4 - Do the "terms of service" of the platforms available in your country authorize the copying and storage for the purpose of constituting "training data" and the creation of "AI outputs" of the works posted by the users of the platform? If so, give examples of the relevant Terms of Service.

There seems to be few examples concerning the platforms operated by the Japanese companies which authorize or prohibit the exploitation works for the purpose of training or the use of AI outputs.

But the 'mimic' (<https://illustmimic.com/en/>) is a platform that accepts solely their users' own artistic creations for the purpose of constituting "training data". The platform provides the 'service that uses AI to automatically create illustration makers that reflect the individuality of the artist', which has the terms of service, under which '[t]he rights to the illustrations created by mimic belong to the artist, not to the management of the platform'.

By the way, there is no provision on the prohibition of a contractual override in the Japanese Copyright Act.

2.5 - Are you aware of the conclusion of individual or collective licenses on this point? If yes, in which fields of creation? Under what conditions? If so, give examples.

No.

3. Using AI as a tool for rights management and administration

3.1 - To what extent is AI used to locate or identify protected content, to moderate it, or even to fight against infringement?

No specific information is available.

3.2 - If computer tools are used for this identification, are there rules to allow the evaluation of the tools used in order to verify the relevance of the results produced by the AI system? (For example, in the framework of the European Digital Services Act, platforms have an obligation of transparency, notably on the tools used and the results they produce - art. 15).

If the answer is yes, are these rules derived from practice (usages, contracts, softlaw...) or imposed by legislation or regulation, or by case law?

There is no specific legal rule for using AI system in Japan. The JDLA "Model Guideline for Using Generative AI" is not likely to be concerned with this point.

3.3 - To what extent is AI used as a tool to recommend protected content? For example, the proposal of "playlists" by Pandora or any other online communication service making recommendations of works.

Regarding the services operated by the Japanese companies, there is no specific information is available.

3.4 - Should we fear, through this recommendation, a risk of dilution of contents and revenues due to a possible opacity of the system?

| *No idea.*

3.5 - Does your national or regional law contain transparency obligations on the use of an AI system for rights management in your national or regional law (e.g. the European Digital Services Act)? What are they?

| *There is no specific obligation for using AI system.*

3.6 - In general, do these tools have to comply with rules in terms of product safety or conformity? Are there procedures for certification of these tools by an authority or by professional associations? Are suppliers subject to specific due diligence obligations?

| *There is no specific rule for AI system in terms of product safety or conformity. There is no procedures for certification of these tools by an authority or by professional associations.*

Artificial intelligence and literary and artistic property

The contours of protection

The status of AI Outputs

1. Access to protection

- Characterization of the AI output as a "Work" of authorship

Note: If an AI output has all the external aspects of a work of authorship, is it possible to consider it as a work of authorship protected by copyright?

4.1 - Does a "Work" always imply the presence of a physical person?

4.2 - From what threshold is it possible to consider that there is a human intervention giving rise to an original work in the realization of an AI output? What types of intervention would allow to know if this threshold has been crossed?

4.3 - How can we distinguish between AI-assisted outputs and outputs generated by an AI?

4.4 - In some countries, it is asserted that there can only be a work of authorship if the form obtained is the result of creative work by the author in the sense that the latter is aware of the result (work) he wants to achieve even if this result is a little different from his hope/expectations. This requirement, for example, would exclude the quality of author of a person deprived of discernment (for example, an insane person, a very young child, a somnambulist...) or would entail the refusal of protection of a production which would be only the fruit of random forces.

Does this condition exist in your country?

If so, is it a statutory or administrative requirement? Does it derive from caselaw? From secondary authorities (e.g. academic writings)?

4.5 - Are the criteria traditionally considered to be irrelevant (such as merit, or purpose) taken into account in the framework of protecting an AI output?

- Characterization of a performer's performance

4.6 - In order to be vested with a neighboring right, does the performer necessarily have to be a natural person?

In other words, is an "interpretation" from an artificial intelligence protectable under neighbouring rights?

4.7 - In order to be vested with a neighbouring right, must the performer necessarily interpret a work created by a natural person?

In other words, is the interpretation, by a human being, of a production of artificial intelligence protectable under neighboring rights? (Suppose an AI-generated musical composition: if performed by a human being, would the performance be protectable?)

- If the AI output does not qualify for copyright protection

4.8 - Are the productions generated by AI, that are not covered by copyright, in the public domain?

4.9 - In your country, could the productions generated by AI be qualified as "commons" (it being understood that, in some countries, the notion of "commons" has a different meaning than "public domain")? Under what conditions or according to what criteria?

4.10 - How can we be sure that the creation presented as realized by an author is not an artificial production?

4.11 - Usually, a collective management organization (CMO) manages a catalog attached to an author without making distinctions between "works" / "productions". How to manage the case of an author whose usual works belong to his repertoire but who would also use an AI system to generate other "productions"?

2. The rights regime

- The choice of the right (nature, ownership, regime, limitations)

** As your legislation currently stands:*

5.1 - Is the output generated by an artificial intelligence system likely to be protected by copyright in your country?

5.2 - If applicable, does the production generated by an artificial intelligence system benefit from a full copyright, in particular as regards the duration and scope of the rights, or from a modified or special right?

5.3 - If there is a protection by an adapted or special copyright (as it exists sometimes for certain works, as for example, in Europe, concerning computer programs), what are the modifications or adaptations?

5.4 - Who is the author? Who would be the owner of the rights? Could the output be considered a joint work? If so, between whom and in what cases?

5.5 - Is there a special ownership rule (presumption, or even fiction, as it exists in some countries for computer-generated creations; see for example, art. 9 (3) Copyright, Designs and Patents Act (CDPA) in England)?

** In the event of a possible legislative change:*

Are there any concrete proposals in your country related to the items listed below? If so, answer questions 5.6 and following.

If not :

i) the national rapporteurs can give their personal opinion while making it clear that these are mere proposals of secondary authorities (e.g., academics) and not positive law;

ii) or they can go directly to the questions numbered 6 and following.

5.6 - What would be the criteria to be retained to allow access to copyright protection for AI outputs?

5.7 - Should a specific copyright be created for these productions?

5.8 - With what particularities (e.g., duration and content of the rights) ?

5.9 - Can there still be a moral right ?

5.10 - Should there be a special ownership rule (presumption, or even fiction, as it exists in some countries for computer-generated creations)?

5.11 - Should a deposit be required? / A declaration of "origin"?

5.12 - Should a kind of neighbouring right or a sui generis right be created?

5.13 - What would be its characteristics?

5.14 - The rights covered?

5.15 - Generally speaking, what would be the limitations on or exceptions to this new right?

5.16 - How should this protection be articulated with other existing protections?

5.17 - In the absence of protection by a property right, are there any compromise solutions?

For example, a kind of paying public domain for them: collection of royalties paid to a collective management organization for distribution among authors continuing to create works in the traditional way? What else?

- AI and violation of rights: the choice of remedy

6.1 - Can an AI output infringe, and to what extent? Who would be liable?

An AI output would infringe copyright if there is reliance and similarity to "work". The user is liable in case of AI-assisted outputs, according to a government report. But discussion continues as to who will be liable, user or system programmer.*

**https://www.kantei.go.jp/jp/singi/titeki2/tyousakai/kensho_hyoka_kikaku/2017/johozai/houkokusho.pdf (Page38)*

The JDLA Model Guideline art. 6 (2) discourages to use AI systems which learned only an author's works. The same article also discourages to input in prompts titles of existing works and names of authors. These conducts are likely to make AI systems infringe copyrights of existing works and authors.

6.2 - Are there other legal means (e.g. unfair competition, parasitism) to engage the liability of the person responsible for the AI output? (Who would that person be?)

In the personal opinion of a National Reporter (NAGATSUKA), Japanese Unfair Competition Prevention Act Article 2(1) may be applicable to whom, par example, run an AI-assisted text generation system, when it is named like an author and generates texts in his or her style after it has been trained by his or her works.*

**Unfair Competition Prevention Act (Act No. 47 of May 19, 1993)(Definitions) Article 2(1)*

The term "unfair competition" as used in this Act means any of the following:

(i)the act of creating confusion with another person's goods or business, by using an indication of goods or business (meaning a name, trade name, trademark, marks, containers or packaging for goods belonging to a person's business, or any other indication of a person's goods or business; the same applies hereinafter) that is identical or similar to another person's indication of goods or

business that is well-known among consumers as belonging to that person, or by transferring, delivering, displaying for the purpose of transfer or delivery, exporting, importing or providing through telecommunications lines goods that use such indication;

6.3 - Beyond copyright, can personality rights prevent the realization by an AI of a production using the voice or physical aspect of another person?

As long as the person is alive, whose personality rights can be exercised to the extent infringed by the AI of a production.

- Question of transparency and remuneration

7.1 - In your country, is there a requirement (legal, administrative, jurisprudential, arising from practice) that AI-generated content in general be declared as such (see for example in Europe, the AI Act of April 21, 2021² and the more nuanced position of the Council of the European Union of November 2022³)?

The JDLA Model Guideline art. 6 (5) encourages to respect ChatGPT Service Policy, which oblige users to declare the use, when the user make available the contents generated by it.

(Optional) If not, do you think that such a solution should be adopted?

7.2 - If applicable, how is the sharing and payment of remuneration carried out when AI is involved in the creative process?

Lack of information on it.

(Optional) If there is no existing solution, what solution do you think should be adopted?

7.3 - If applicable, how is the sum linked to the AI allocated (cultural action? payment to other rights holders...)?

Lack of information on it.

(Optional) If there is no existing solution, what solution do you think should be adopted?

National Reporters

UENO Tatsuhiro (part2 and 3) and NAGATSUKA Makoto (part 1, 6 and 7) supported by JASRAC and Geidankyo CPRA

For further information, we attach afterwards [Tatsuhiro UENO "Copyright Issues on Artificial Intelligence and Machine Learning" \(2017\)](#), already available at <http://www.f.waseda.jp/uenot/Copyright-AI-IJCAI2017.pdf>. Attention, this article was written before the enactment of article 30-4, actual copyright limitation for machine-learning.

Presentation of the article 30-4 in French by NAGATSUKA is included in the book "L'entreprise et l'intelligence artificielle - Les réponses du droit" (EAN : 9782361702618) edited by A. Mendoza-Caminade in 2022⁴.

² <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A52021PC0206>

³ <https://www.consilium.europa.eu/fr/press/press-releases/2022/12/06/artificial-intelligence-act-council-calls-for-promoting-safe-ai-that-respects-fundamental-rights/>

⁴ <https://www.lgdj.fr/l-entreprise-et-l-intelligence-artificielle-9782361702618.html>

Copyright Issues on Artificial Intelligence and Machine Learning

Tatsuhiko UENO

Professor of Intellectual Property Law at Waseda University
Deputy director of the Research Center for the Legal System of Intellectual Property (RCLIP)
Tokyo, JAPAN
uenot@waseda.jp

Abstract

Artificial intelligence (AI) has recently been causing intellectual property (IP) issues including copyrightability and legal protection of AI-generated works, pre-trained models and training data sets.

AI-generated works are not eligible for copyright protection due to lack of human author's intellectual creation, in most countries including Japan. Against this background, it is now under discussion in Japan whether it is necessary to introduce some sort of legal protection for AI/computer-generated works for the purpose of protecting investment for them.

Also, pre-trained models and training data sets can be protected by copyright, as long as they are creative (or original) and considered as not AI-generated works but human author's own intellectual creation.

Additionally, it should be noted that the Japanese Copyright Act has the explicit provision on copyright exception for text and data mining (Art.47*septies*) under which it shall be allowed to copy any copyright-protected work for the purpose of machine learning without authorization of copyright holders. This provision is quite helpful for machine learning and facilitating or boosting technological development of AI, since it applies to machine learning not only for a non-commercial purpose but also for a commercial purpose. Hence, Japan is the paradise for machine learning.

1 Introduction

The development of artificial intelligence (AI) has recently been causing IP (intellectual property) issues, especially copyright issues.

In Japan, they are current hottest subjects of discussions, which include copyrightability of AI-generated works, pre-trained models as well as training data sets and other legal protection of them, for the purpose of facilitating technological development of machine learning and promoting (re)utilization of them

Additionally, it should be emphasized that the Japanese Copyright Act has the explicit provision on copyright exception for text and data mining including machine learning even for commercial purpose (Art.47*septies*), which is quite helpful for facilitate technological development of AI.

The purpose of this article is to overview the copyright issues on AI by focusing on the recent discussion in Japan.

2 Copyright protection of AI-generated works

The first problem is whether AI-generated works (e.g. music,¹ painting,² novel³) are eligible for copyright protection or not.

2.1 Discussions since 1972 in Japan

It has been discussed in Japan since 1972, whether computer-generated works are eligible for copyright protection.

The second subcommittee (for considering computer issues) of the Council for Copyright Law was established in 1972, which examined problems over computer including computer-generated works. The final report was published in June 1973.⁴

¹ See e.g. "Illiac Suite" (String Quartet No. 4) in 1957 by Lejaren Hiller and Leonard Isaacson using "ILLIAC I" computer, the music album titled "Bach by Design: Computer Composed Music" by David Cope (1994), the song titled "Daddy's Car" (2016) after the style of Beatles (Sony CSL Research Laboratory) and the AI systems "Orpheus" (University of Tokyo, et al.) <<http://www.orpheus-music.org/v3/>> as well as "Iamus" (University of Málaga).

² See e.g. Next Rembrandt Project <<https://www.nextrembrandt.com/>>.

³ See e.g. The AI project of Prof. Hitoshi Matsubara at Future University Hakodate et al. <http://www.fun.ac.jp/~kimagure_ai/>.

⁴ The final report (in Japanese) is available at <http://www.cric.or.jp/db/report/s48_6/s48_6_main.html>.

The ninth subcommittee (for considering computer-generated works issues) of the Council for Copyright Law was established in 1985 and the final report was published in 1993.⁵ According to the report, computer-generated works can be copyrightable in some cases. In such cases, a person who contributes to works (operator of computer system) would be considered the author of the work in such cases.

Recently, due to the advancement of AI technology, copyright issues on AI are current hottest subjects of discussions in Japan and there are a number of committees or meetings established by the Japanese Government on this issue since 2015, most of which I am a member.⁶

In June 2015, Intellectual Property Strategy Headquarters of Cabinet Secretariat of Japan published the Intellectual Property Strategy Program 2015.⁷ Then, the Committee for Considering the Next Generation IP System discussed this issue from November 2015 to April 2016.⁸ Subsequently, the Committee on New Information Goods discussed this issue from October 2016 to March 2017.⁹

2.2 Japan: No copyright protection for computer-generated works

In general, copyright is a right which protects a work as an author's own intellectual creation. That leads there must be a human author for copyright to subsist in a work. Since a computer cannot be an author and AI-generated work cannot be considered a copyright-protected work in the meaning of copyright law, even though such AI-generated work might be entirely new and highly artistic.

In most countries including Japan, there is no special provision on AI/computer-generated work and it is generally considered that there is no author in AI-generated works that are not considered copyright-protected works.

Actually, it is true that a work created by human author using a computer as a tool is considered a copyright-protected work.

However, it is often difficult to distinguish between a work created by a human author using a computer as a tool and AI-generated work.

2.3 UK: Copyright protection for computer-generated works

On the other hand, under UK Copyright Act (CDPA), a "computer-generated" work, namely a work which "is generated by computer in circumstances such that there is no human author of the work" (Sec.178), is also eligible for copyright protection, the "copyright expires at the end of the period of 50 years from the end of the calendar year in which the work was made" (Sec.12(7)) and "the author shall be taken to be the person by whom the arrangements necessary for the creation of the work are undertaken" (Sec.9(3)). The right to be identified as author (Sec.77) and the right to object to derogatory treatment of work (Sec.80) do not apply in relation to computer-generated work (Sec.79(2)(c) and Sec.81(2)).¹⁰

Similar provisions can be seen in Ireland (Arts.21(f) and 30), New Zealand (Arts.5(2)(a) and 22(2)), India (Art.2(d)(6)), South Afrika (Art.1(1)(5)), Barbados (Art.10(4)) and Hongkong (Arts.11(3), 17(6), 91(2)(c), 93(2) and 198).

2.4 Is legal protection needed?

If AI-generated works are not eligible for copyright protection and totally free from copyright, it causes a problem whether some sort of legal protection for AI-generated work is necessary or not.

Discussions are still now ongoing in Japan. There seems to be a consensus that copyright protection is too much and excessive for the AI-generated works, even if it might be necessary to introduce some sort of legal protection for AI-generated works for the purpose of protecting substantial investment for them.

Therefore, there are some opinions to introduce a legal protection other than copyright and weaker than copyright protection.

But the difficult problem is that it is almost impossible to recognize whether a work is created by a human being or generated by a computer, even if you might seriously observe it. Therefore, if the legal protection for AI-generated works is weaker than copyright protection, no one reveal the fact that the work was generated by AI.

⁵ The final report (in Japanese) is available at <http://www.cric.or.jp/db/report/h5_11_2/h5_11_2_main.html>.

⁶ Also, regarding patent law issues on AI, the committee on creations using AI and 3D printing data was established at Japan Patent Office and Institute of Intellectual Property in September 2016. The final report (in Japanese) was published in 2017 which is available at <https://www.jpo.go.jp/shiryou/toushin/chousa/pdf/zaisanken/2016_10.pdf>.

⁷ The translation of the Intellectual Property Strategy Program 2015 is available at <http://www.kantei.go.jp/jp/singi/titeki2/kettei/chizaikeikaku2015_e.pdf>.

⁸ The final report (in Japanese) is available at <http://www.kantei.go.jp/jp/singi/titeki2/tyousakai/kensho_hyoka_kikaku/2016/jisedai_tizai/hokokusho.pdf>.

⁹ The final report (in Japanese) is available at <http://www.kantei.go.jp/jp/singi/titeki2/tyousakai/kensho_hyoka_kikaku/2017/johozai/houkokusho.pdf>. The translation of the overview of the final report is available at <http://www.kantei.go.jp/jp/singi/titeki2/tyousakai/kensho_hyoka_kikaku/2017/johozai/houkokusho_e.pdf>.

¹⁰ See Thomas Margoni / Mark Perry, From Music Tracks to Google Maps: Who Owns Computer Generated Works?, 26 Computer Law and Security Review 621, 2010; Andres Guadamuz, Do androids dream of electric copyright? Comparative Analysis of Originality in Artificial Intelligence Generated Works, 2017(2) Intellectual Property Quarterly 169.

Thus, some scholars propose to introduce copyright protection for AI/computer-generated works like in UK.¹¹

Also, there is an opinion to impose criminal penalties on a person who pretends that he created AI-generated work by himself in public.¹²

3 Legal protection of pre-trained models

The second problem is whether pre-trained models are eligible for copyright protection or not.

Although it is unclear whether a pre-trained model is considered as a copyrightable work, pre-trained models can be basically protected by copyright as works of computer program, as long as they are creative (or original) and considered as not AI-generated works but human author's own intellectual creation.

Similarly, in case you fine-tune a copyright-protected pre-trained model A in a creative (or original) way, the fine-tuned pre-trained model B can be eligible for copyright protection as a derivative work.

If you copy such a copyright-protected pre-trained model without authorization of the copyright holder, it can constitute copyright infringement.

However, it should be noted that "ideas" underlying in copyright-protected works are generally not eligible for copyright protection, since copyright does not protect ideas but expressions (called "Idea-Expression Dichotomy"). Hence, so-called "distillation" of a pre-trained model A does not constitute copyright infringement, even if the newly created pre-trained model B has the same function as pre-existing pre-trained model A, because expressions of the pre-existing pre-trained model A are not directly copied in the course of distillation.

Instead, ideas can be protected by patent rights. Generally speaking, ideas with regard to pre-trained models can be patentable, as an invention of a product or a process (including a process of producing a product). As a result, if a distillation of a pre-trained model A is regarded as producing of patented invention of the pre-trained model A and if the newly created pre-trained model B has the same function as pre-existing pre-trained model A, it constitutes an infringement of the patent right.

As I mentioned before, if a pre-trained model is generated by a computer, it is neither protected by copyright nor patent right, since there must be a human author or inventor for copyright or patent protection. On the other hand, if a pre-trained model is created by human being using a computer as a tool, it can be eligible for copyright and patent protection.

Moreover, it should be noted that pre-trained models are basically protected as trade secret under unfair competition law or other relevant law, as long as they are kept secret and not publicly known.

However, the problem under recent discussion in Japan is whether some sort of legal protection for pre-trained models is needed or not for the purpose of protecting substantial

investment, since forming pre-trained models often involves a great cost to analyze a tremendous amount of data by using supercomputers.

Quite recently, the Japanese Government formed the Committee at Ministry of Economy, Trade and Industry in 2017 to consider whether to introduce new sort of legal protection for valuable big data which is under a certain technological control for the purpose of facilitating transactions or reusing of big data by amending the Unfair Competition Prevention Act, even if it is not kept secret or is publicly known and therefore is not considered trade secret. Such big data might include pre-trained models. Therefore, the proposed amendment would facilitate reuse or sharing of pre-trained models.

4 Legal protection of training data sets

The third problem is whether training data sets are eligible for copyright protection or not.

Training data sets can be protected by copyright as works of compilation. If you copy such a copyright-protected training data sets without authorization of the copyright holder, it constitutes copyright infringement.

However, creativity (or originality) is generally required for copyright protection. A training data set in which selection and arrangement of data are not creative (or original) is not eligible for copyright protection.

Thus, it is recently also debated whether some sort of legal protection for a valuable training dataset enabling effective machine learning is needed or not, because it enables an effective machine learning and has a great economic value.

5 Copyright exception for facilitating machine learning

5.1 Style is free from copyright as idea

If you input a number of copyright-protected works by a particular author (e.g. all songs of the Beatles) to a computer, conduct machine learning and generate a pre-trained model which can generate "new" works after the author's (e.g. Beatles) style, such newly generated works do not infringe copyright of the author, since copyright does not protect "style" as idea, as mentioned above ("Idea-Expression Dichotomy").

5.2 Japan as paradise for machine learning?

Machine learning analyzes a tremendous amount of data and sometimes requires copying of copyright-protected works to be analyzed. In general, it is necessary to obtain authorization from copyright holders to copy copyright-protected works.

¹¹ See Yoshiyuki Tamura, Chosakukenhō Gaisetsu [Outline of the Copyright Act], (2nd ed., 2001) p.243 (in Japanese); Koji Okumura, Jinkōchinō ga umidasita contents to chosakuken [AI-

generated contents and copyright], 70-2 Patent 15 et seq. (2017) (in Japanese).

¹² See Okumura *supra* note 11 at 15 et seq.

However, the Japanese Copyright Act¹³ has already introduced the explicit provision on copyright exception for data analysis or text-and-data mining (Art.47*septies*) in 2009,¹⁴ under which the storage in a media for the purpose of a statistical analysis by using a computer shall not constitute copyright infringement.

As a result, you are allowed to copy all of the “Star Wars” movies, all of the Beatles’ albums or all of the Japanese manga or animations without any authorization of copyright holders, if you conduct it for machine learning in Japan.

In Europe, the European Commission published the Proposal for a Directive of the European Parliament and of the Council on Copyright in the Digital Single Market on September 14, 2016, which has the provision on copyright limitation for text and data mining (Art.3),¹⁵ under which “reproductions and extractions made by research organisations in order to carry out text and data mining of works or other subject-matter to which they have lawful access for the purposes of scientific research” would be allowed.

UK has already introduced Section 29A of the UK Copyright Act (CDPA) [Copies for text and data analysis for non-commercial research] in 2014. However, this provision only applies to the computational analysis “for the sole purpose of research for a non-commercial purpose” (sec.29A).

On the other hand, Article 47*septies* of the Japanese Copyright Act applies to the computational analysis not only for a non-commercial purpose but also for a commercial purpose as well. Therefore, this explicit provision must be quite helpful for machine learning and facilitating or boosting technological development of AI.

Furthermore, the Japanese Copyright Act also introduced Article 30*quater* [Exploitation for the use in a test for the development or the practical use of technology] in 2012, which could also apply to exploitations of works for the use in a test for the development of AI technology.

I would say that Japan is the paradise for machine learning in the world.

5.3 Further development in Japan

In Japan, it is planned to expand the scope of Article 47*septies* of the Japanese Copyright Act in order to respond machine learning. There seems to be a good possibility that the Japanese Government will submit the bill to amend the Japanese Copyright Act to the next 194th extraordinary Diet session which will seemingly start in autumn of 2017.

If the amendment will come into force, a wide variety of machine learning would be explicitly allowed, including a

provision of training data sets to other persons conducting machine learning or a joint project of machine learning by a number of companies, which would facilitate reuse of training data sets.

6 Conclusion

AI is recently attracting global attention in the field of law and it is now the hottest topic in Japan. To be honest, I am skeptical about whether AI has a vast impact on copyright law. However, in the face of the rapid development of AI technology, it would be meaningful to consider copyright issues on AI, not only copyrightability of AI-generated works, pre-trained models and training data sets but also copyright exception for machine learning.

However, the problem we have to ask here is how to strike a balance between protection and free (re)use of pre-trained models, training data sets and AI-generated works, for the purpose of facilitating technological development of machine learning, promoting utilization as well as reuse of them.

¹³ Translations of the Japanese Copyright Act are available at <<http://www.cric.or.jp/english/clj/>> and <<http://www.japaneselawtranslation.go.jp/law/detail/?id=1980&vm=2&re=&new=1>>. Regarding the outline written in English of the JCA 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).

¹⁴ The Act for Partial Revision of the Copyright Act, Act No.53 of August 19, 2009. 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.

¹⁵ Proposal for a Directive of the European Parliament and of the Council on Copyright in the Digital Single Market, 2016/0280 (COD), <<http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52016PC0593&from=EN>>.