

# 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,<sup>1</sup> painting,<sup>2</sup> novel<sup>3</sup>) 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.<sup>4</sup>

---

<sup>1</sup> 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).

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

<sup>3</sup> See e.g. The AI project of Prof. Hitoshi Matsubara at Future University Hakodate et al. <[http://www.fun.ac.jp/~kimagure\\_ai/](http://www.fun.ac.jp/~kimagure_ai/)>.

<sup>4</sup> The final report (in Japanese) is available at <[http://www.cric.or.jp/db/report/s48\\_6/s48\\_6\\_main.html](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.<sup>5</sup> 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.<sup>6</sup>

In June 2015, Intellectual Property Strategy Headquarters of Cabinet Secretariat of Japan published the Intellectual Property Strategy Program 2015.<sup>7</sup> Then, the Committee for Considering the Next Generation IP System discussed this issue from November 2015 to April 2016.<sup>8</sup> Subsequently, the Committee on New Information Goods discussed this issue from October 2016 to March 2017.<sup>9</sup>

## 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)).<sup>10</sup>

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.

---

<sup>5</sup> The final report (in Japanese) is available at <[http://www.cric.or.jp/db/report/h5\\_11\\_2/h5\\_11\\_2\\_main.html](http://www.cric.or.jp/db/report/h5_11_2/h5_11_2_main.html)>.

<sup>6</sup> 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/shiryu/toushin/chousa/pdf/zaisanken/2016\\_10.pdf](https://www.jpo.go.jp/shiryu/toushin/chousa/pdf/zaisanken/2016_10.pdf)>.

<sup>7</sup> The translation of the Intellectual Property Strategy Program 2015 is available at <[http://www.kantei.go.jp/jp/singi/titeki2/kettei/chizaikeikaku2015\\_e.pdf](http://www.kantei.go.jp/jp/singi/titeki2/kettei/chizaikeikaku2015_e.pdf)>.

<sup>8</sup> 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](http://www.kantei.go.jp/jp/singi/titeki2/tyousakai/kensho_hyoka_kikaku/2016/jisedai_tizai/hokokusho.pdf)>.

<sup>9</sup> The final report (in Japanese) is available at <[http://www.kantei.go.jp/jp/singi/titeki2/tyousakai/kensho\\_hyoka\\_kikaku/2017/johozai/houkokusho.pdf](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](http://www.kantei.go.jp/jp/singi/titeki2/tyousakai/kensho_hyoka_kikaku/2017/johozai/houkokusho_e.pdf)>.

<sup>10</sup> 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.<sup>11</sup>

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

### 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.

<sup>11</sup> See Yoshiyuki Tamrura, 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).

<sup>12</sup> See Okumura *supra* note 11 at 15 et seq.

However, the Japanese Copyright Act<sup>13</sup> has already introduced the explicit provision on copyright exception for data analysis or text-and-data mining (Art.47*septies*) in 2009,<sup>14</sup> 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),<sup>15</sup> 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 194<sup>th</sup> 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.

---

<sup>13</sup> 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).

<sup>14</sup> 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.

<sup>15</sup> 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>>.