



文化庁

Agency for Cultural Affairs,
Government of Japan

"General Understanding on AI and Copyright in Japan" -Overview-

(published by the Legal Subcommittee under the Copyright Subdivision of the Cultural Council)

Japan Copyright Office (JCO)
(Copyright Division, Agency for Cultural Affairs, Japan)
May 2024

- This document is an overview of the "General Understanding on AI and Copyright in Japan" (hereinafter, the "**General Understanding**") published by the Legal Subcommittee under the Copyright Subdivision of the Cultural Council. For detailed information, please refer to the main document.
- The General Understanding represents the subcommittee's views on the interpretation of the current Japanese copyright act as of the time of publication. The General Understanding is not legally binding, nor should it be considered as a definite legal assessment of any specific generative AI or related technologies that currently exist.

Basic Concepts on AI and Copyright

Basic Concepts of the Japanese Copyright Act

- ❑ The Japanese Copyright Act aims to strike **a balance between protecting the rights and interests of copyright holders and ensuring the smooth exploitation of copyrighted works.**
- ❑ This concept is also crucial when considering the matters on AI and copyright.

Background of the Study by the Council

- ❑ With the rapid development and widespread use of generative AI in recent years, there are concerns about the lack of clarity regarding the relationship between AI and copyright.
- ❑ Normally, the provisions of the Copyright Act should be interpreted by the judiciary on a case-by-case basis. However, there are currently very few court precedents that directly address the relationship between AI and copyright.
- ❑ In order to address these concerns without waiting for court precedents to accumulate, since July 2023, a national council* composed of experts has discussed how the current Copyright Act should be applied in relation to AI aiming to provide a general understanding of this topic.

*Legal System Subcommittee under the Copyright Subdivision of the Cultural Council (consisting of law scholars, lawyers, judges and other experts with knowledge of copyright law and other intellectual property laws)

Three Perspectives on AI and Copyright

- When it comes to AI and copyright, it is important to differentiate between the exploitation of copyrighted works in the "AI development / training stage" and infringement in the "generation / utilization stage".

AI Development / Training Stage

- Creating a training data set by collecting and reproducing copyrighted works as training data
- Developing AI (pre-trained machine learning model) by using such data sets for training

Generation / Utilization Stage

- Generating materials (e.g., images etc.) using AI
- Uploading and making public AI-generated images, selling copies of generated images (e.g., publishing art collection books, etc.)

- Moreover, it is also necessary to separately consider whether AI-generated materials (i.e., content generated by AI) are susceptible to copyright protection and can become "copyrighted works".

Main Topics Discussed by the Council

- The Council discussed the relationship between AI and copyright, focusing on the concerns of stakeholders such as AI-related businesses, users, creators, and other right holders. The main topics covered during the discussion included the following:

AI Development / Training Stage

Exploitation of copyrighted works for AI development / training etc. (Article 30-4 etc.)

- The “non-enjoyment purpose” requirement
- Article 30-4 proviso

Generation / Utilization Stage

Copyright infringement in the generation and utilization of AI-generated materials

- Criteria for copyright infringement
- Countermeasures for copyright infringement
- Cases where AI-related businesses are liable for copyright infringement

Copyrightability of AI-generated Material

Criteria for determining the copyrightability of AI-generated material

- The relationship between the specificity of instructions to the AI and the copyrightability of AI-generated materials

Overview of the General Understanding

: AI Development / Training Stage

Article 30-4 of the Copyright Act

- ❑ Exploitation of a copyrighted work not for enjoyment of the thoughts or sentiments expressed in the copyrighted work (exploitation for non-enjoyment purposes)* such as AI development or other forms of data analysis may, in principle, be allowed without the permission of the copyright holder. *e.g., collection (i.e. reproduction) of copyrighted works as AI training data
- ❑ “Enjoyment” under the Article 30-4 refers to the act of obtaining the benefit of having the viewer’s intellectual and emotional needs satisfied through using the copyrighted work.

《Examples of acts that can be called “enjoyment”》

Literary works : To read

Musical works

: To appreciate

Works of computer programming : To execute

Movie works

- ❑ The financial benefits that copyright holders receive from their works are generally considered rewards for meeting intellectual and emotional needs. Meanwhile, the exploitation of works for non-enjoyment purposes, which may occur without the consent of the copyright holder, is generally regarded as not harming the financial interests of the copyright holder. Therefore, in such cases acquiring permission for use of the copyrighted works from the copyright holder is not deemed to be required pursuant to Article 30-4 of the Act.

Article 30-4 of the Copyright Act

- ❑ The provisions of Article 30-4 of the Act do not apply to the "exploitation of works for the purpose of enjoyment" and the "exploitation of works where the main purpose is non-enjoyment such as provision for use in data analysis, but where there is also the purpose of enjoyment.*"

* The presence or absence of an "enjoyment" purpose is determined by "the work" exploited under the Article, not by other copyrighted works or not copyrighted elements such as an "artist's style".

- ❑ Where a work is being used for "non-commercial" or "research purposes", etc., permission from the copyright holder is required where there is also a "purpose of enjoyment" of the work present.
- ❑ Furthermore, Article 30-4 of the Act does not apply in "cases that would unreasonably prejudice the interests of the copyright holder.*"
 - e.g., reproducing a copyrighted database work for the purpose of data analysis, such as AI training for which licenses for data analysis are available in the marketplace, etc.

Article 30-4 of the Copyright Act (Act No. 48 of 1970)

(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:

- i. if it is done for use in testing to develop or put into practical use technology that is connected with the recording of sounds or visuals of a work or other such exploitation;
- ii. 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; the same applies in Article 47-5, paragraph (1), item ii);
- iii. if it is exploited in the course of computer data processing or otherwise exploited in a way that does not involve what is expressed in the work being perceived by the human senses (for works of computer programming, such exploitation excludes the execution of the work on a computer), beyond as set forth in the preceding two items.

Cases not Meeting the "Non-enjoyment Purpose" Requirement

- ❑ In the following cases, the reproduction of copyrighted works for AI training does not satisfy the "non-enjoyment purpose" requirement since the "purpose of enjoyment" is also present, rendering Article 30-4 of the Act inapplicable.

Exploitation of Works for AI Training (e.g., Collection of AI Training Data)

- ✓ The collection of works for AI training to generate materials similar to copyrighted works within the collected works. This is a form of additional training on the foundation model, also known as fine-tuning, and intentional "overfitting".

Exploitation of Works in Situations other than AI Training

- ✓ The collection of works as input data to generative AI for implementation of retrieval augmented generation (RAG)*

*Those that aim to output the creative expression (wholly or partially) of a copyrighted work inputted.

* Even if Article 30-4 of the Act does not apply, the exploitation of copyrighted works for output by RAG, etc. (and reproduction or public transmission of copyrighted works as a preparatory act for RAG, etc.) may be made without the permission of the right holder as long as the requirements of Article 47-5 of the Act are.

Overview of the General Understanding

: AI Development / Training Stage

Imitation of a "Creator's Style"

- ❑ Copyright protection applies to the "creative expressions" of an idea and not to the "idea" itself. Therefore, an AI-generated material which applies the "creator's style" of a preexisting copyrighted work does not infringe copyright if the style merely encompasses an idea.
- ❑ Conversely, Article 30-4 of the Act may not apply to AI training which "picks off" a specific creator's style in the following situations:

- The distinction between "ideas" and "creative expressions" has been determined on a case-by-case basis, depending on the specific circumstances.
- A small group of works consisting only of specific creator's copyrighted works may not only have a common "creator's style" (i.e., idea), but may also include common "creative expressions."
- If such a group of works is intentionally reproduced in order to carry out additional training for the purpose of generating materials that contains all or parts of such common creative expressions, Article 30-4 of the Act does not apply to such reproduction because the "purpose of enjoyment" coexists.

* In addition, if AI-generated material applies creative expressions with such group of works described above at the generation / utilization stage, this may constitute copyright infringement.

Overview of the General Understanding

: AI Development / Training Stage

Article 30-4 Proviso: if unreasonably prejudice the interests of the copyright owner

- ❑ It is necessary to assess the applicability of the Article 30-4 proviso by considering “whether it will compete in the market with the copyrighted work” and “whether it will impede the potential sales channels of the copyrighted work in the future.” This assessment should be done by taking various factors into account, such as “technological advancements” and “changes in the way the copyrighted work is used.”^{*1}
- ❑ The following explains the relationship between the collection of AI training data and this proviso:

- The proviso applies where a copyrighted database work which contains data provided in a format that can be used for information analysis and is available on the internet for a fee was reproduced without compensation.
- In certain cases, it is assumed that there is an intention to commercialize copyrighted database works including website data, that are organized in a format suitable for information analysis. This assumption^{*2} arises when, for example, “technical measures”^{*3} have been taken to prevent reproduction of copyrighted works for AI training.”
- In cases where copyrighted database works^{*4} are reproduced for AI training while avoiding above mentioned technical measures such reproduction falls within the Article 30-4 proviso.

*1 From [“Basic concepts regarding flexible provisions on the restriction of rights that respond to advancement in digitalization and networking”](#) (PDF, in Japanese), JCO.

*2 In addition to the below-mentioned “technical measures,” the past sales record of database works that can be used for information analysis, etc.

*3 Measures to prevent web crawlers from collecting AI training data include adding instructions in a file called “robots.txt” in the website's root directory or requiring access authentication through ID and password.

*4 Possible methods of reproducing website content include using web crawlers to collect large amounts of data.

Reproducing Infringing Copies for AI Training

- ❑ Collecting AI training data from websites while being aware that they distribute pirated or copyright-infringing content should be strictly avoided.
- ❑ If a business (i.e., AI developer or AI service provider) knowingly collects AI training data that includes infringing copies, it may be held liable for copyright infringement by the generative AI in the following cases:

- According to legal precedent, a person other than the direct infringer may be liable for copyright infringement.
- If an AI developer or AI service provider collects training data for their AI from a website that they know contains pirated or infringing content, there is a high possibility that the business will be held responsible for any copyright infringement caused by the generative AI developed using the training data taken from the website.

* According to the “General Understanding”, it is recommended for copyright holders to share information regarding websites that contain pirated or other infringing copies with the businesses. This way businesses can identify these websites and exclude them as a source for data collection for AI training data.

Requirements for Copyright Infringement: "Similarity" and "Dependence"

- ❑ When AI-generated images or copies of thereof are uploaded to social media or sold, copyright infringement will be determined based on the same criteria as for normal infringement.
- ❑ In other words, if an AI-generated image or any other creation is found to have similarity (i.e., common creative expression) and dependence (i.e., creation based on existing copyrighted work) with an existing image, etc. (copyrighted work), and there are no applicable copyright exceptions, it will be considered an infringement of copyright.
- ❑ If an AI product does not have "similarity" or "dependence" with a preexisting copyrighted work, it does not constitute copyright infringement of the copyright in that work. According to the Copyright Act, it then is possible to use the copyrighted work without the permission of the copyright holder. Additionally, permission is not required for uses falling within copyright exceptions, such as the act of creating and viewing images for personal use (i.e., reproduction for private use).

Overview of the General Understanding

: Generation / Utilization Stage

“Dependency” in the Case of Generation by AI

- ❑ The requirements for finding copyright infringement include the elements “similarity” and “dependency.” In the case of AI generated content dependency can be explained as follows:

When It Is Unknown whether the Existing Copyrighted Work is Used for AI Training

- If it is uncertain whether a particular copyrighted material is used in the AI training data, dependency will be presumed if the copyright holder can prove that “the AI user had access to the existing copyrighted work” or “the AI-generated material has a high degree of similarity with the work”. This means that it is possible for the copyright holder to establish dependency even though it is unknown whether the AI has used the work during AI training.

When the Copyright Holder can Prove the Existing Copyrighted Work is Used for AI Training

- It is generally assumed that there was dependency on a preexisting copyrighted work, even if the user of an AI was not aware of it*, if the work was used for AI training during the development stage of that AI.

* However, if the generative AI “does not output the creative expression of the copyrighted work used for AI training at the generation / utilization stage”, it may be judged that there is no dependency.

- ❑ It is important to remember that while AI can legally “generate materials”, caution should be exercised when “using the generated materials”, such as uploading them to social media platforms. Just because the materials were generated legally does not necessarily mean they can be used legally in all situations and the legality* of each specific situation needs to be assessed and determined. * e.g., whether or not copyright exceptions apply

Overview of the General Understanding

: Generation / Utilization Stage

Countermeasures for Copyright Infringement

□ In case of copyright infringement, the copyright holder of existing works can take following countermeasures:^{*1}

- The copyright holder may seek injunctive relief against the party that infringes on their copyright by using generative AI. This injunction may include “prohibiting the creation of new infringing material” as well as the “use of infringed works already created.” Additionally, the copyright holder may request the destruction of AI-generated materials that were created as a result of the infringement.
- In certain cases,^{*2} the copyright holder may request that the infringing material be removed from the AI training dataset used for development of the generative AI that created it.
- Furthermore, in certain circumstances,^{*3} the copyright holder may request from the AI developer and/or the AI service provider that appropriate measures^{*4} are taken to prevent copyright infringement caused by the generative AI.

*1 The exact measures that may be taken in response to copyright claims will be determined by a court, considering their necessity and other relevant factors.

*2 I.e., if an AI developer is responsible for copyright infringement related to the creation and application of their product, and there is a high probability that the dataset utilized for the development of the generative AI which produced the infringing product could also be used for AI development in the future.

*3 I.e., in cases where an AI developer or AI service provider is liable for copyright infringement arising from the generation and use of a product and where there is a high probability that further infringing material will be generated by the generative AI that produced the infringing product.

*4 For instance, there are certain ways to limit the technical capabilities of generative AI. This includes: 1.) implementing measures to prevent the AI from generating outputs based on certain inputs or 2.) inhibiting the creation of similar materials that may infringe on copyrighted works used for the AI's training.

Cases in which AI-related businesses are liable for copyright infringement

- ❑ Based on current judicial precedents, it is possible for both AI users and businesses that develop or provide AI services to be held liable for copyright infringement related to AI-generated materials. Liability for copyright infringement will depend on various factors.

Factors that Increase the Likelihood that Businesses will be Held Liable

- ✓ A particular generative AI tends to produce materials that infringe on copyright laws frequently.
- ✓ In developing and providing generative AI, the business has not taken measures to prevent the generative AI from generating copyright infringing materials, despite recognizing that it is highly likely to do so.

Factors that Reduce the Likelihood that Businesses will be Held Liable

- ✓ The business has taken measures to prevent the generative AI from generating copyright infringing materials in developing and providing generative AI.

* In addition, if a generative AI is not designed to produce copyrighted materials at a high rate, but such materials are created due to malicious instructions from a user, it is unlikely that the business will be held liable for infringement. This, however, only applies where the generative AI is in a state where it does not generate infringing materials at a high rate.

Overview of the General Understanding

: Copyrightability of AI-generated Material

Copyrightability of AI-generated Material

- According to the Copyright Act, a (copyrighted) “work” is defined as a “creatively produced expression of thoughts or sentiments that falls within the literary, academic, artistic, or musical domain.”
- Materials autonomously generated by AI* are not “creatively produced expressions of thoughts or sentiments” and are therefore not considered (copyrighted) “works.”

*i.e., material that is generated by AI without any instructions from humans (or only by giving simple instructions as prompt (e.g., “Draw a cat.”))

- On the other hand, if AI is used as a “tool” by a person to creatively express thoughts or sentiments, such material is considered a “work”, and the user of the AI the “author”.

* Please note that only a person (i.e., a natural or legal person) can be an “author” under the Copyright Act. An AI itself, which does not have a legal personality, cannot be an author.

- Determining whether a person has used AI as a “tool” depends on two factors: whether the person had a “creative intention” and whether the person has made a “creative contribution”.

Overview of the General Understanding

: Copyrightability of AI-generated Material

Relationship between the specificity of instructions given to an AI and the copyrightability of AI-generated materials

- The following outlines under which circumstances AI products are recognized as containing the AI user's "creative contributions."^{*1}

- The copyrightability of AI-generated content will be determined on a case-by-case basis.^{*2}
- Specifically, it is assessed through a comprehensive consideration of the extent of creative contributions that surpass mere effort.

Examples of factor in determining copyrightability of AI-generated material

- ✓ Amount of instructions/input: "Detailed instructions that specifically indicate what constitutes creative expression" are more likely to be considered as creative contributions. However, lengthy instructions (i.e., prompts) that merely suggest an idea do not influence the assessment of creative contribution.
- ✓ Number of generation attempts: A large number of attempts alone does not affect the assessment of creative contribution. Repeated attempts, while checking the generated materials and correcting the instructions/input, may be recognized as a creative contribution.
- ✓ Selection from multiple output materials: The mere act of selection itself does not influence the determination of creative contribution. However, certain elements of choice may be involved which may be considered as creative.
- In addition, any additions or corrections made by humans to AI-generated materials that can be considered creative expressions are generally considered to be copyrighted works.

^{*1} The issues of "whether an AI-generated work can be copyrighted or not" and "whether generation by AI or the usage of AI-generated materials constitutes the copyright infringement" need to be separately. Even if an AI-generated material is copyrighted, it can still constitute copyright infringement of the existing copyrighted work, if it meets the necessary requirements for infringement.

^{*2} Hence, whether or not an AI product corresponds to a copyrighted work is not solely determined by "the fact that it is an AI product".

Continued Information Gathering and Discussion

- ❑ The General Understanding is a summary of the Council's tentative consensus about AI and copyright. The Council will continue to gather information on topics such as the "collection of legal precedents on AI and copyright," the "development of AI and related technologies," the "progress of research on AI and copyright in other countries," and more. As necessary, the Council will consider revising the General Understanding.
- ❑ JCO will actively continue to collect examples of AI copyright infringement cases (including suspected cases) through the consultation desks for the public.*
- ❑ The discussion within the General Understanding mainly focused on the relationship between AI and copyright (i.e., property rights). However, it is important to consider if there are any other aspects that need to be taken into account independently. It is also necessary to explore how AI intersects with moral and related rights, including the use of performances/records that feature the voices of actors/voice actors.

※ The consultation desks for the public set up by the Agency for Cultural Affairs, Japan are as follows:

- Consultation desk for countermeasures against copyright infringement due to pirated copies sold on the Internet
<https://www.bunka.go.jp/seisaku/chosakuken/kaizoku/index.html>
- Legal consultation desk regarding cultural and artistic activities
https://www.bunka.go.jp/seisaku/bunka_gyosei/kibankyoka/madoguchi/index.html

Effective communication between relevant parties

- ❑ To ensure the implementation of good practices in the intersection of AI and copyright, it is important to facilitate not only legal discussions but also effective communication among relevant parties involved in AI and copyright matters, such as copyright holders (i.e. creators), AI developers, AI service providers, and AI users.
- ❑ Generative AI technology is thus far built upon the creative activities of humankind. If the sustainability of human creativity is compromised, the sustainable development of generative AI technology becomes unattainable. Hence, it is hoped that a mutually beneficial relationship fostering a fruitful content creation environment and cultural development will be established between generative AI and its associated businesses, as well as creators.
- ❑ Establishing appropriate rules and guidelines between involved parties based on current legal interpretations will be essential as a foundation for future discussions. This entails fostering a common understanding among the relevant parties, particularly regarding a correct understanding of generative AI and related technologies and mechanisms, as well as the sentiments of creators and other copyright holders concerning the use of their copyrighted works.
- ❑ JCO, in collaboration with other relevant ministries and agencies, will continue its efforts to facilitate effective communication among the involved parties and to promote best practices in the intersection of AI and copyright.