GENERATIVE AI FROM A PERSPECTIVE OF COPYRIGHT LAW — WHAT IS PERMITTED AND WHAT IS PROBLEMATIC —

Yui Matsuura
Intellectual Property Dept., Technology & Innovation Studies Div.
Mitsui & Co. Global Strategic Studies Institute

SUMMARY

- The use of generative AI involves risks related to intellectual property laws, including those for copyrights. This report summarizes the basic concepts of generative AI and copyright law, differences in regulations among countries, and risks and future challenges.
- The relationship between generative AI and copyright law can be considered divided into two phases: the "development and learning phase" and the "generation and utilization phase." In Japan, there is a high degree of freedom in the use of data for learning purposes, and the country is sometimes referred to as a "machine learning paradise." However, new debates are emerging regarding the balance between the use of data by generative AI and the protection of rights.
- Copyright law does not protect AI-generated content as copyrighted works unless a creative contribution by humans is recognized. On the other hand, issues such as copyright infringement by AI-generated content and the need for protection have also been raised.

1. WHERE ARE THE RISKS IN THE GROWING USE OF GENERATIVE AI?

BASIC CONSIDERATIONS IN GENERATIVE AI AND COPYRIGHT LAW

The rapid advancement of AI technology in recent years has led to greater use of generative AI, and it is now being utilized in various ways in business. At the same time, the use of generative AI and content created by generative AI (hereinafter referred to as "AI-generated content") can raise legal issues, especially under copyright and other intellectual property laws.

The relationship between generative AI and copyright law can be understood in two phases: the "development and learning phase" and the "generation and utilization phase." Those are thought to have the following respective risks.

- **Development and learning phase:** Risks related to the legality of training data
- **Generation and utilization phase:** Risks related to copyright infringement by AI-generated content, protection of AI-generated content, and unauthorized use of such content

Since copyright law adopts the principle of territoriality,[1] the applicable laws vary depending on the country where the development/training, generation/use, etc., take place. And since the risks of copyright infringement

[1] This refers to the principle that a country's copyright law applies only within its own borders. In other words, Japanese copyright law applies only in Japan, and US copyright law applies only in the US.

and the feasibility of commercial use can vary significantly by country, the rules of each country need to be confirmed.

2. GENERATIVE AI AND COPYRIGHT LAW: JAPAN

2-1. CONCEPTS IN THE DEVELOPMENT AND LEARNING PHASE

Generative AI is trained using large amounts of data, and if that data includes copyrighted works, there is a potential risk of copyright infringement. However, under Japan's Copyright Act, such use is permitted as long as it is not for the "purpose of enjoying the work[2]" (Article 30-4(ii)), thereby granting a high degree of freedom for data use for learning purposes. Because of this, Japan is sometimes referred to as a "paradise for machine learning.[3]" This environment not only helps accelerate technological development for AI developers, but also promotes the advancement and practical application of AI models, influencing a wide range of fields such as healthcare, robotics, and creative industries. It is also attractive to overseas companies, encouraging them to enter the Japanese market and establish research facilities.

2-2. CONCEPTS IN THE GENERATION AND UTILIZATION PHASE

Copyright law protects works created by humans. **AI-generated content is not considered a copyrighted work unless there is a recognized creative contribution by humans[4].** At the same time, AI-generated content may infringe existing works.

The criteria for judging copyright infringement apply equally to AI-generated content and existing works, with the requirement of "Dependence[5]" and "Similarity." For example, if AI-generated content was an illustration that resembles the character "Doraemon," and the user entered "Doraemon" as a prompt, dependence would likely be found. Importantly, even if the user did not intend to copy, the presence of "Doraemon" illustrations in the training data could still lead to a finding of dependence.

Additionally, the issue of misrepresenting AI-generated content as one's own work[6] has been raised. This could enable the exercise of rights that would otherwise not be recognized and raise concerns about insufficient protection for those who invested in the work believing it to be original. There have been discussions regarding the need for a new protective framework for AI-generated content, which is not currently protected under existing copyright law, and future developments will be closely watched.

2-3. TERRITORIALITY AND APPLICABLE LAW

With generative AI services, it is clear that Japan's Copyright Act applies when the entire process from development to provision and use is completed within Japan. On the other hand, if a Japanese business trains generative AI on servers located in Country X and uses that AI to provide services to users in Japan, the question arises as to which country's copyright laws apply—Japan's or Country X's.

In recent years, **there has been a trend toward a more flexible interpretation of the principle of territoriality** in cases of copyright infringement via the Internet. Based on this premise, this case can be summarized as follows.

---

2 Behavior that satisfies intellectual and emotional desires, such as reading literary works or enjoying music and movies, falls under "purpose of enjoying."

3 Example: Tatsuhiro Ueno, "Data Mining and Copyright: Japan as 'Paradise for Machine Learning'" *Journal of the Japanese Society for Artificial Intelligence*, 36 (6) (Nov. 2021): page 745-749

4 The act of humans making direct contributions to creative works

5 Creating works by referencing existing works

6 The issue of misattributed content

- **Development and learning phase:** If learning work is substantially carried out in Japan, Japan is deemed to be the place of use, and Japan's Copyright Act is applied.
- **Generation and utilization phase:** If AI-generated content is generated in Japan, Japan's Copyright Act is applied.

3. GENERATIVE AI AND COPYRIGHT LAW: SYSTEMS IN INDIVIDUAL COUNTRIES AND GLOBAL ISSUES

3-1. US

The US has adopted a legal system based on common law,[7] which has features that differ from Japan's Copyright Act. In both the development and learning phase and the generation and utilization phase, there are no specific provisions for the use of existing works, and whether or not fair use[8] applies becomes the point of contention in copyright infringement disputes.

The US has the world's largest generative AI market, with approximately 70% of the world's major AI models developed by companies and academic institutions based in the US.[9] Meanwhile, several lawsuits have been filed alleging copyright infringement in connection with training of generative AI. The development of legal frameworks and regulations remains an ongoing issue.

3-2. EUROPE

In Europe, the EU AI Act, a regulatory framework for AI, has been adopted. It includes provisions on intellectual property rights, including copyright. Member states may introduce additional regulations tailored to their national circumstances, based on the following common rules.

Common rules (EU AI Act)

- Transparency obligations: Developers of generative AI are obligated to publish summaries of training data.
- Copyright compliance: Developers are obligated to exclude content from AI training if the copyright holder has opted out.

For example, in the UK, third-party copyrighted works may not be used for AI training **for commercial purposes** without permission. Unlike Japan, this approach places greater priority on protecting copyright holders and is far from a "paradise for machine learning."

3-3. CHINA

China's copyright law does not contain clear provisions regarding the copyrightability of AI-generated content or the learning behavior of generative AI, leaving those to be governed by other laws.

In February 2024, a court in China ruled for the first time recognizing the liability of a generative AI service provider for damages. A Chinese licensee of the "Ultraman" science fiction series sued for copyright infringement, alleging that images were used as training data without permission, and the court ruled that the unauthorized use constituted infringement. This ruling suggests that **explicit permission is required** for the

7 Common law is a legal system based on case law, while Japan's legal system is a codified civil law system.
8 This is a legal principle that allows the reuse of works without the copyright holder's permission if it falls under the category of fair use.
9 How CEOs Can Navigate the New Geopolitics of GenAI | BCG

use of copyrighted works for training purposes in China, indicating that the concept of fair use is narrower than in Japan and the US.

In January 2025, it was reported that Chinese AI company DeepSeek may have used the output of OpenAI's large-scale model to build a smaller model through "distillation.[10]" Whether this method constitutes copyright infringement is legally unclear, and it is difficult to determine under current copyright laws.

Figure: Summary of Systems in Individual Countries

Phase	Japan	US	Europe	China
Development / Training	"Paradise for Machine Learning" where there is a high degree of freedom in the use of data.	The applicability of fair use is at issue.	There is an obligation of transparency, and opt-out is possible.	Explicit permission is required for data use.
Generation / Utilization	[Common] AI-generated content that does not involve creative contribution by humans is not protected as copyrighted works, and the criteria for judging copyright infringement are the same as for ordinary works.			
		The applicability of fair use is at issue.	There is an obligation to clearly indicate that it is AI-generated content.	There are no clear provisions, but it has been demonstrated in court precedents.

Source: Compiled by MGSSI

4. ISSUES IN DATA UTILIZATION AND RIGHTS PROTECTION FOR GENERATIVE AI

4-1. USERS APPREHENSIVE OF THEIR CONTENT BEING USED FOR GENERATIVE AI TRAINING

In November 2024, social media platform X (formerly Twitter) explicitly stated that it would utilize the content of users' posts for AI training. While the content may be used for training of "Grok" conversational AI which is developed by X, no opt-out option is provided to completely exclude one's content from being used for that training.

This decision sparked a strong backlash from creators in particular. If content is used for AI training without the creator's consent, AI-generated content could compete with the original and potentially displace it in the market. Some creators have taken measures such as adding watermarks stating "No Unauthorized AI Learning" to their works, but these are not legally binding and have limited effectiveness.

The evolution of generative AI is supported by vast amounts of data. However, while development companies gain the benefits of that, the users who provide the data receive no compensation or rights in return. Under this imbalanced profit structure, major tech companies could monopolize the AI market, potentially putting creators and general users at a disadvantage. Establishing a framework that provides fair recognition and protection to users who contribute to the advancement of AI will be a major issue moving forward.

4-2. DISCUSSIONS IN JAPAN

In many countries, establishing appropriate frameworks for recognizing and compensating data providers is seen as an important issue, but the development of specific systems and legal frameworks is still in its early stages. In Japan as well, discussions on this issue are active, and how to balance the use of data by generative

[10] A method of transferring the knowledge (essence learned once) held by a large model to a small model

AI with the protection of rights is a major point of discussion for the “Study Group on Intellectual Property Rights in the AI Era” hosted by the Cabinet Office.

As previously mentioned, the “development and training phase” and the “generation and utilization phase” need to be considered separately in discussions on generative AI. In Japan, only the “development and training phase” is currently a “machine learning paradise,” and copyright infringement in the “generation and utilization phase” is not permitted. Therefore, appropriate measures to prevent infringing content from being generated (output) will also be required.

5. THE FUTURE OF GENERATIVE AI AND LEGAL REGULATIONS: PROSPECTS FOR INTERNATIONAL COORDINATION

Copyright law is being forced to adapt further with the technical advancement and widespread adoption of generative AI. Currently, legal risks vary between countries, making it essential to establish operational frameworks that comply with regulations in the regions where generative AI is used. Going forward, the following issues will likely become key points of discussion: (1) revising the scope of copyright law to address issues such as the legality of training data and the protection of AI-generated content, (2) establishing international rules, and (3) balancing interests with those of data providers. Given that generative AI is used across national borders, the establishment of international rules is particularly desirable.

In order to minimize legal risks associated with generative AI and maximize business opportunities, understanding of domestic and international legal frameworks and swift response are needed.