

Some Legal Issues Concerning the Use of Copyrighted Works in the Development and Learning Phases of Generative AI



Kei Teshirogi
kei.teshirogi@ohebashi.com

I. Introduction

In recent years, generative AI has rapidly advanced and become widespread. However, legal interpretations and practical implementations related to this field have struggled to keep pace with its rapid evolution. Moreover, significant differences in opinion have been arising from the different perspectives of those creating generative AI, those utilizing it, and those whose copyrighted works are being used by generative AI.

This article will introduce the status of discussions in Japan on two major issues. It will first discuss the relationship between the use of copyrighted works in the development and learning phases of generative AI and Article 30-4 of the Copyright Act of Japan, which is one of the many issues concerning generative AI in Japan.

The Copyright Act was amended in 2008 to establish a provision like Article 30-4 thereof, which limits the rights of copyright holders. This is rare even from a comparative legal viewpoint. The said article was established as a limitation on rights provision to flexibly respond to new forms of use of copyrighted works in accordance with technological innovation. Its interpretation has become a point of contention, particularly concerning the use of copyrighted works during the development and learning stages of generative AI.

Additionally, this article will cover the debate over the “override problem,” which is a question of whether acts that are lawful under copyright law can be restricted by terms of use and similar agreements.

II. Article 30-4 of the Copyright Act

1. An Overview of Article 30-4 of the Copyright Act

Article 30-4 of the Copyright Act stipulates that “[I]t 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.”

The economic value of a copyrighted work lies in its utility whereby those who view such work can enjoy the ideas or emotions expressed therein, thereby satisfying their intellectual and spiritual needs. The Copyright Act is designed to protect such economic value.

Article 30-4 specifies that the act of utilizing a work for a purpose other than such enjoyment (i.e., “**use for a non-enjoyment purpose**”) would not constitute



copyright infringement because it does not typically infringe the interests of the copyright holder. Article 30-4(ii) provides that “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)” is a typical example of use for a non-enjoyment purpose.

During the development and training stages of generative AI, copyrighted works may be replicated as training data. This action involves “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,” and therefore constitutes “use in data analysis” as described in Article 30-4 (ii). Consequently, the use of copyrighted works for data analysis during the development and training stages of generative AI is, in principle, lawful under Article 30-4.

However, a major point of contention in the interpretation of Article 30-4 is whether the said article applies to the use of copyrighted works as training data for generative AI if such use is both for a non-enjoyment purpose, such as data analysis, and an enjoyment purpose.

2. Application of Article 30-4 to Acts where an Enjoyment Purpose and a Non-enjoyment Purpose Coexist

What constitutes use where both enjoyment and non-enjoyment purposes coexist? A typical example of use with such coexisting purposes is the act of using a large number of images of a specific character (such as Mickey Mouse) in a training dataset to develop a specialized image generation AI that only generates characters similar to that particular character. Such action would involve using data for training for data analysis, which is a non-enjoyment purpose, while simultaneously aiming to adjust the

trained parameters so that the essential features of Mickey Mouse are reproduced as training data, which constitutes an enjoyment purpose.

The prevailing view in Japanese practice is that Article 30-4 would not apply when the two purposes coexist. In other words, for additional training of an existing trained model (including the collection and processing of training data for this purpose), if it is conducted with the intent of producing an output of a creative expression of the copyrighted work contained in the training data, then Article 30-4 would not apply and such additional training would constitute copyright infringement.

In contrast to the above view, it has been opined that even for such usage, the act of using copyrighted works in the development and training stages of generative AI is still a data analysis activity and it would be difficult to conclude that the provision would not apply based on its wording. However, this opinion also holds that specialized AI should be regulated under the proviso of Article 30-4 and thus, the conclusion that the use of copyrighted works for developing specialized AI constitutes copyright infringement would remain unchanged.

3. Interpretation of the Proviso of Article 30-4 of the Copyright Act

Article 30-4 of the Copyright Act, while making the use for non-enjoyment purpose lawful, provides in the proviso 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.*”

Whether a use unreasonably harms the interests of the copyright holder is determined from the perspective of whether it conflicts with the copyright holder’s market for the copyrighted work or obstructs potential future markets for the work. For example, the use of a



database created for the purpose of providing data analysis falls under this proviso because such use would conflict with and unjustly harm the copyright holder's interests in the market for training data.

Aside from such example, determining other cases that might be regulated by the proviso requires further discussion. Considering the similarity between Article 30-4 and the fair use doctrine in the U.S., it is useful to consider the fair use criteria in the interpretation of the said proviso, including: (a) the purpose and character of the use, (b) the nature of the copyrighted work, (c) the amount and substantiality of the portion used, and (d) the effect of the use on the market for the copyrighted work.

III. The “Override Problem”

1. What is the “Override Problem”?

On internet sites that provide copyrighted works, there are cases where the terms of service restrict usage actions for data analysis, which is permitted under Article 30-4. The issue at hand is whether such terms of service can limit the lawful actions under Article 30-4. This is referred to as the “override problem” (hereafter, such restrictive clauses will be referred to as “**override clauses**”). The so-called “override problem” can be divided into two issues: (a) whether the terms of service containing override clauses constitute a contract in the first place, and (b) if so, whether such override clauses are enforceable.

2. Formation of Contracts with Override Clauses

Whenever the content of an internet site is used during the training phase of a generative AI, there are two scenarios where the terms of service with override clauses can form part of the agreement between the content provider and the user, namely: (a) when the terms of service are deemed standard terms under Article 548-2, paragraph 1 of the Civil Code, thereby forming part of the agreement by default, and

(b) when, although the terms do not qualify as standard terms, an offer and acceptance of the individual terms of the agreement have been established. Since the terms of service in question are likely to be considered standard terms under the Civil Code, the author will examine below the validity of an agreement formed based on standard terms.

According to Article 548-2, paragraph 1 of the Civil Code, the terms of service shall be considered standard terms if (a) there is an agreement on the standard transactions (transactions conducted with an unspecified number of parties where all or part of the terms are standard and reasonable for both parties), and (b) either (i) an agreement exists to include standard terms as part of the contract, or (ii) the preparer of the standard terms had previously indicated to the other party that the terms would be included in the contract. In such cases, the individual clauses of the standard terms would also be deemed part of the contract.

Regarding the requirement in item (a) in the paragraph above, transactions on internet sites that provide content for users to download are often considered standard transactions due to the unspecified number of users and the uniform conditions of such transactions. Although there is a debate about when there exists a standard transaction agreement, given that the significance of the Civil Code provisions lies in acknowledging the binding nature of contracts even if concluded with somewhat minimal intent in the case of standard terms, it is possible to view the act of downloading as evidence of a user's agreement to the standard transaction.

As to the requirement in item (b) of the paragraph mentioned above, in the context of internet transactions, the applicability of sub-item (ii) is crucial, and whether such indication of the inclusion of such standard terms is present in an internet transaction



must be examined on a case-by-case basis. However, generally speaking, it is often not particularly difficult for content providers to structure internet sites in a way that would ensure that users encounter the terms of service before completing a transaction (such as downloading). Therefore, simply listing the terms on pages or locations that users might not necessarily visit before the transaction may not meet the requirement of sub-item (ii).

Even if the requirements for implied agreements under Article 548-2, paragraph 1 of the Civil Code are met, paragraph 2 thereof states that *“notwithstanding the provisions of the preceding paragraph, the person is deemed not to have agreed to any provisions as referred to in that paragraph that restrict the rights or expand the duties of the counterparty and that are found, in light of the manner and circumstances of the standard transaction as well as the common sense in the transaction, to unilaterally prejudice the interests of the counterparty in violation of the fundamental principle prescribed in Article 1, Paragraph 2.”*¹

This provision regulates unfair or unexpected clauses. Override clauses, which restrict lawful usage actions that users would normally be able to perform, and which are generally not recognized by users as restrictions on lawful use, are seen as having a significant element of unexpectedness.

Therefore, even if the terms of service meet the requirements of Article 548-2, paragraph 1 of the Civil Code, there is a substantial possibility that override clauses may not be deemed contracts under paragraph 2 thereof.

3. Validity of Override Clauses

Even if terms of service with override clauses are deemed contracts, whether such override clauses that restrict lawful actions are valid is a separate issue. If Article 30-4 of the Copyright Act were interpreted as a mandatory provision, then override clauses that contradict it would be deemed invalid without needing to consider the specific elements of the case. However, since limitations on rights under copyright law are generally interpreted as discretionary provisions, the purpose of the override clauses and other factors must be comprehensively considered to determine whether such override clauses are invalid for being contrary to public order and morals (Article 90 of the Civil Code).

In this regard, the Study Group on New Intellectual Property System Issues² issued a report that identified several factors to consider, namely, the purpose of the limitation on rights provisions, the extent of the disadvantage thereof to users, copyright holders and providers, and the circumstances related to good faith and fairness between the parties. Given that Article 30-4 of the Copyright Act also aims to achieve public interest objectives, such as promoting AI innovation and facilitating the smooth use of information, recognizing the validity of override clauses could potentially undermine the purpose of this provision. The report therefore concluded that there is a significant possibility that such clauses could be deemed invalid for being contrary to public order and morals. The report, however, also emphasized the need to consider the circumstances of each case, noting that it is difficult to predict the specific disadvantages to copyright holders or their ripple effects from using the internet content. It is also possible that, based on the context leading up to the transaction, there may be cases where, from the perspective of fairness, protecting user interests is less

1. Article 1(2) of the Civil Code sets forth the general principle of good faith and stipulates that “[T]he exercise of rights and performance of duties must be done in good faith.”

2. Study Group on New Intellectual Property System Issues, “Report” [Houkokusyo] (2022) (available at https://www.meti.go.jp/policy/economy/chizai/chiteki/pdf/reiwa3_itaku_designbrand.pdf#page=143) (in Japanese).



critical. Consequently, at present, the validity of override clauses must be determined on a case-by-case basis. The accumulation of future cases for further clarification is being awaited.

VI. Conclusion

This paper introduced the legal issues concerning the use of copyrighted works in the development and training phases of generative AI, particularly focusing on the interpretation of Article 30-4 of the Copyright Act and the “override problem.” As mentioned above, the opinions on these issues are not yet settled and some aspects thereof remain unclear. Given the need for active legal discussions regarding generative AI, it is the hope of the author that this article will contribute to such ongoing dialogue.

DISCLAIMER

The contents of this Newsletter are intended to provide general information only, based on data available as of the date of writing. They are not offered as advice on any particular matter, whether legal or otherwise, and should not be taken as such. The authors and Oh-Ebashi LPC & Partners expressly disclaim all liability to any person in respect of the consequences of anything done or omitted to be done wholly or partly in reliance upon the whole or any part of the contents of this Newsletter. No reader should act or refrain from acting on the basis of any matter contained in this Newsletter without seeking specific professional advice.