18 April 2024 | News | Future of IP | Yoshinori Okamoto And Jin Yoshikawa

Gorodenkoff / Shutterstock.com

Is Japan still a machine learning paradise?

Despite inevitable ambiguity, published analysis on Japanese copyright law and AI will provide important insight until cases are decided in the courts, say Yoshiniori Okamoto of Yuasa and Hara and Jin Yoshikawa of Adams and Reese.

The capabilities of artificial intelligence (AI) in recent years have stunned the world. Their pace of growth is even more shocking, with advanced multimodal generative AI, AI agents, and glimmers of artificial general intelligence peeking over the horizon. Clarifying the copyright implications of AI is therefore an urgent imperative for legal systems around the globe.

In Japan, a unique statutory provision was added to its Copyright Act in 2018, which earned Japan the moniker of "machine learning paradise" by Facebook's chief AI scientist Yann LeCun. With the recent technological developments in generative AI, this provision has become a flashpoint between the interests of AI developers and content creators.

Thus in 2023, a committee within Japan's Agency for Cultural Affairs began penning a document to clarify the interplay of Japan's copyright law and AI, collecting the inputs of lawyers, creators, engineers, and corporate stakeholders across industries and disciplines. Nearly 25,000 public comments were submitted to the drafting committee. And on March 15, 2024, the final 45-page document (the "Document") was published.

https://www.worldipreview.com/future-of-ip/is-japan-still-a-machine-learning-paradise

Though the Document is officially non-binding and does not provide definitive legal evaluations of specific AI services available today, it will provide important insight for practitioners until actual AI copyright cases are adjudicated in Japanese courts.

This short article only covers the highlights of the Document, with a focus on its discussion of Article 30-4 of the Copyright Act. The ambiguity that this Document left regarding some legal questions has arguably chilled some of the optimism of AI developers and, at the same time, done little to quell the concerns of creators. Nevertheless, the Document will help analyse AI copyright cases in the future.

The basic principles of Japanese copyright law

The Document begins with a recap of the basic principles or premises of Japanese copyright law.

As in most modern copyright regimes, Japanese copyright law attempts to balance the interests of authors and the public. A "work" is defined by the Copyright Act as a (1) creatively produced (2) expression of (3) thoughts or sentiments (4) that falls within the literary, academic, artistic, or musical domain. Japanese copyright law adopts the "expression-idea" dichotomy, meaning ideas (including styles of works) are not protected. In practice, identifying whether a particular thing is an expression or an idea is not easy.

Under Japanese law, copyright is a limited bundle of enumerated rights, such as right of reproduction and right to transmit to the public.

Japanese law does not have a "fair use" defence, like the US. Instead, it defines statutory exceptions— eg, reproduction for private use (Article 30), quotation (Article 32-1), reproduction in schools and other educational institutions (Article 35). The document discusses Article 30-4 and Article 47-5 (minor exploitation incidental to data processing and the provision of the results thereof) in more detail.

One last preliminary issue as it pertains to AI copyright cases is choice of law. The Document states that this will be a case-by-case judgment by the courts, but mentions some factors that would increase the chances Japanese law will apply to an AI case: (1) the data collecting program for machine learning is hosted on a server in Japan and reproduces copyrighted works, (2) the generative AI is hosted on a server in Japan and generates copyrighted works, and (3) the generative AI service on the internet publicly transmits content including copyrighted works to users in Japan.

Article 30-4 and the learning phase of generative AI

In many countries, analysis of AI and copyright is typically split into two stages: (1) the learning phase and (2) the generation phase. This split assumes present-day generative AI technologies such as text, audio, or video generative AIs, so it may not be the right framework for AI systems of the future. For now, the Document follows this split to analyse AI and copyright issues in Japan today.

The learning phase of generative AI can raise copyright issues because the collection, preprocessing, and use of training data can directly involve or result in the reproduction of copyrighted works.

In 2018, the Japanese Copyright Act added Article 30-4, which generally excepts from copyright protection the use of work for "non-enjoyment purposes"（非享受目的). This means a copyrighted work may be used so long as the use is not for the purpose of enjoying the work oneself or allowing others to enjoy the work.

For instance, copying and feeding a painting into a machine learning algorithm as RGB data is often a use not for the purpose of enjoying the artwork but for data analysis. It is not surprising that the provision has garnered the attention of AI developers worldwide and become so controversial among creators who publish works in Japan.

However, there are several aspects of Article 30-4 that require closer analysis.

First, what happens if an AI service has more than one purpose? The Document offers the interpretation that, to take advantage of the non-enjoyment exception, there must be no enjoyment purpose whatsoever.

On one hand, this interpretation appears to significantly curtail the exception, as even a minor intent to allow others to enjoy the work would exclude AI services with primarily legitimate purposes from its ambit. On the other hand, the exception may still be easy to satisfy, as it may be unusual that an AI model was trained with the specific purpose of enjoyment of the works that were used in training, rather than enjoyment of the AI outputs. (The Document also notes that, in some cases, Article 47-5 will apply when enjoyment purposes exist.)

For example, the Document suggests that the non-enjoyment exception may not apply if an AI service provider deliberately fine tunes a model in order to produce certain copyrighted works verbatim for its users. This calls to mind the **New York Times**' allegations against OpenAI of "memorisation" of articles. However, if "overfitting" of works occurs unintentionally, the Document suggests the exception may apply depending on the facts, including intent presumed from the AI outputs, as well as the prompts that were used to generate the allegedly infringing outputs. The Document is not legally binding, and it will be up to the court judgment.

Under the same analysis, the Document discusses that the collection and use of a database of works for Retrieval Augmented Generation (RAG) may also be entitled to the non-enjoyment exception if the conditions are satisfied.

Next, the Document explores an exception to the non-enjoyment exception in Article 30-4 for uses that "unreasonably prejudice the interests of the copyright holder, considering the type and purpose of the work and the manner of use". As this "unreasonably prejudice" language is ambiguous, many creators hoped this language would protect them.

Conversely, AI developers hoped for clarifications that would cabin the chilling effect of this proviso. The resulting Document pays respect to both sides without taking a firm stance.

The Document begins that the "unreasonably prejudice" proviso is a "highly flexible" analysis by design, given the technology and the valuable new uses of copyrighted works are rapidly developing in the market.

Then the Document discusses some of the most anticipated battlegrounds.

- Does the mass production of works that are similar to the ideas or styles of a particular creator and that diminish the demand for the creator's work constitute unreasonable prejudice? The Document explains both sides. But it notes that even if copyright law does not provide the desired remedy, creators might have remedies for economic injuries outside of copyright law.

- Does the use of a proprietary database (of, say, news articles) to train a model constitute unreasonable prejudice? Potentially yes. The Document offers a few potentially aggravating factors, such as the AI developer's failure to pay the data provider for access and the data provider's offering of a licence specifically for data mining. A mitigating factor may be the data provider's failure to employ technical measures to prohibit data mining or web crawling, such as a robots.txt file. Going forward, practitioners should pay attention to the legal implications of technological measures to protect works from machine learning.

- Does training on copies of pirated content constitute unreasonable prejudice? The Document emphasises the harm of piracy on creators, although it acknowledges the practical difficulty of identifying pirated content online. The Document urges AI developers to also give due consideration to avoid assisting the proliferation of pirated content.

Finally, the Document discusses the potential injunctive remedies for "unlawful" machine learning. Generally, a model cannot be ordered to be destroyed, but the Document discusses there may be a case, for example, if a model produces infringing works with high probability, where a request for destruction of the model is required.

When is AI output copyright infringement?

The non-enjoyment exception does not cover the generation and utilisation stage of generative AI. In generative AI cases, the Document suggests examining whether there exists substantial similarity (類似性) and reliance (依拠性), like all copyright infringement cases, on a case-by-case basis.

The Document discusses one novel AI-related issue on the reliance element of copyright infringement:

If an AI user is unaware of an existing copyrighted work (and its expressive content), and the generative AI has not been trained on that copyrighted work during its development and learning phase, then even if the AI generates something similar to that copyrighted work, this would be considered a coincidental match. Therefore, reliance is not recognised, and copyright infringement does not occur.

If this interpretation is adopted by the courts, compared to traditional defendants, it is potentially easier to deny reliance by looking at the training data. This would emphasise the importance of creating a legally "clean" data set for training generative AI models.

For a discussion on a proposed solution for collecting the necessary training data and compensating creators called Data Income, a legal regime that compensates creators for contributing machine learning data, see Yoshinori Okamoto, *Intellectual Property Protection about Learning Data for Artificial Intelligence*, Patent, Vol. 70, No. 10 pp. 91–96 (2017), and Yoshinori Okamoto, *Intellectual Property and Artificial General Intelligence*, 8th AGI Study Group, No. SIG-AGI-008-09, JSAI (2018).

Finally, the Document discusses questions about who would be held responsible for such infringement— whether it be the users of generative AI, the developers of the AI, or the entities providing services using the AI. The Document discusses that an AI developer that made reasonable efforts to prevent infringement may be shielded from liability from a user's deliberate attempt to prompt an infringing output.

Is AI output copyrightable?

The Document discusses various other issues, notably, the copyrightability of generative AI outputs. An AI system may not be an author of a copyrightable work in Japan because it is not a legal person. Whether copyright can extend to an AI output prompted by a human author depends on the specificity of instructions given to generative AI and whether the output meets the elements of a copyrightable "work". Even if an AI output is deemed not copyrightable, other causes of action and remedies might be available under tort law.

Conclusion

This Document left some AI and copyright questions unresolved, causing frustration for both creators and AI developers. The reaction is unwarranted, however, because balancing the interests of creators and AI developers is inherently challenging. In fact, some questions were answered, and in some cases, the Document provided AI service providers with actionable advice on good practices to reduce risk.

That said, Japan is probably not a "machine learning paradise" as LeCun originally declared.
For now, it is likely the current Japanese copyright regime in Japan still aspires to balance the interests of creators and AI. In the future, bold solutions such as a Data Income system are likely needed to avoid vague and unpredictable balancing tests and promote the sustainable development of socially beneficial AI.

https://www.worldipreview.com/future-of-ip/is-japan-still-a-machine-learning-paradise

We welcome questions about the Document.

Yoshinori Okamoto **is an attorney at Yuasa and Hara. He can be contacted at** *yokamoto@yuasa-hara.co.jp*

Jin Yoshikawa **is the artificial intelligence team leader at Adams and Reese. He can be contacted at** *jin.yoshikawa@arlaw.com*

https://www.worldipreview.com/future-of-ip/is-japan-still-a-machine-learning-paradise