



Our Thoughts on Training Data and Copyright

Ryuhei Murakami
JIPA Frontier technology study team



About JIPA

- JIPA is a non-profit, non-governmental organization with 1,361 member companies.
- JIPA represents industries and users of the intellectual property (IP) system, and provides related institutions all around the world with well-timed, suitable opinions on improvement of their IP systems and their utilization.
- JIPA contributes to its member companies' business opportunities and promote the sound progress of technology and development of industries.



Our Thoughts on Generative AI

- (1) Regarding machine learning for generative AI, Japan's copyright law is not that "different" or "paradise" compared to other countries**
- (2) When considering the applicability of copyright exception for the use of copyrighted materials in machine learning, the most important question is whether the learning was performed with the "purpose" of outputting a similar work. We should be careful in inferring "purpose" from "results".**
- (3) Soft law measures are also important for opt-out problem and machine learning from copyright-infringing materials.**



Our Thoughts on Generative AI

(1) Regarding machine learning for generative AI, Japan's copyright law is not that "different" or "paradisaal" compared to other countries

- Copyright law of Japan stipulates that copyrighted works can be used without the consent of the right holder "as long as it is used for **"non-enjoyment purpose"** and does not **unreasonably prejudice the interests of the copyright owner**. And the exploitation of training data for AI machine learning is considered a typical example of non-enjoyment purpose use.
- Copyright law of Japan is sometimes referred to as a "machine learning paradise," but even in machine learning, only if there is even **a slight "purpose of enjoyment" at the same time, permission from the right holder is required**. In addition, it should be noted that permission is required for any use that unreasonably harms the interests of the copyright owner. In that sense, Copyright law of Japan is not that "different" or "paradisaal" compared to other countries regarding generative AI.



Our Thoughts on Generative AI

(2) When considering the applicability of copyright exception for the use of copyrighted materials in machine learning, the most important question is whether the learning was performed with the “purpose” of outputting a similar work. We should be careful in inferring “purpose” from “results”.

- Recent deliberations within the Japanese government council (re)confirmed that the "non-enjoyment purpose" in machine learning is the absence of the **“purpose”** of expressing the substantial similarity of training data in the output.
- However, in reality, whether or not there is a "non-enjoyment **purpose**" is inferred based on the **“result”** of to what extent the AI developed based on the training data generates the contents which are substantially similar to the works contained in the training data.
- Although we do not deny that such **“inference”** is necessary for the finding of such **“purpose”**, we argue **that such inference should be made with caution**, and the key factors of such inference should be clarified given that there are also AIs such as the foundation model which are developed without a specific purpose. We have expressed our that opinion in public comments to the government recently.
- We think it is possible that similar arguments can be applied to other countries.



Our Thoughts on Generative AI

(3) Soft law measures are also important for opt-out problem and machine learning from copyright-infringing materials.

- Based on recent deliberations within the Japanese government council, the council expressed its view that exploiting training data from the internet by ignoring opt-out mechanisms such as robots.txt in machine learning, as well as exploiting training data containing copyright infringing materials, is not necessarily and directly considered to “unreasonably prejudice the interests of the copyright owner”. There are opinions from the user side that this conclusion is a reasonable interpretation given the purpose of the legislation and the way of the provision is written, while there are also opposing opinion from right holder side.
- In addition to the above situation, there are also no established rules and practices for rights holders to opt-out (opt-out measure using robots.txt also has its limits) , and there are some cases where it is difficult to distinguish between infringing content and non-infringing content. We argue that it is possible to take **soft law measures** such as issuing international guidelines to **keep balance between not causing chilling effect to users/developers and respecting the will of rights holders**. This argument can be applied to other countries.

Thank you for your attention!

Creating IP Vision for the World



Japan Intellectual Property Association

