

## **Artificial Intelligence – Copyright and the Text and Data Mining Exception**

### **About AVIA**

The Asia Video Industry Association (AVIA) is the trade association for the video industry and ecosystem in Asia Pacific. It serves to make the video industry stronger and healthier through promoting the common interests of its members. AVIA is the interlocutor for the industry with governments across the region, leads the fight against video piracy through its Coalition Against Piracy (CAP) and provides insight into the video industry through reports and conferences aimed to support a vibrant video industry.

### **Text and Data Mining exception**

The rapid growth of AI in all aspects of the video industry (and indeed in all industries) brings lots of challenges, especially in relation to copyright. This has caused regulators around the globe to consider the question of intellectual property protection, resulting in a number of different approaches, introduced in the past 5 years. The specific aim of this paper is to consider Hong Kong's proposed approach in introducing a legal Text and Data Mining (TDM) exception in relation to AI and to examine the challenges of such an approach.

### **Is a TDM exception necessary to protect copyright?**

AVIA does not believe that a TDM exception is necessary. Copyright Law is dynamic, technologically neutral and has been sufficiently flexible to accommodate a range of technological developments since its origin. It is a robust framework and is already equipped with tools to deal with new issues that may arise from using new technologies, including AI.

As new issues arise, HK courts are well placed to approach these questions in a thoughtful and careful manner using the existing framework under HK copyright law.

To the extent developers wish for certainty in their use of copyrighted works to train AI systems, they can always seek licenses from the copyright holders. Licensing markets for training AI models are continuing to develop. Such licensing will also help avoid other concerns about AI use, because the training inputs are likely to be of higher quality, tailored to the specific AI system's needs, and supported by the requisite consents, minimizing risks of error, bias, and violations of privacy.

### **What is HK IPD proposing**

The Hong Kong Intellectual Property Department (IPD) have proposed introducing an amendment to the Hong Kong Copyright Ordinance which would give a TDM exception. This is

being introduced with a “three-step test” (in line with international standards), along with various guardrails including:

- Users must have lawful access to copyright works
- Infringing copies must not be used
- Copyright users are required to keep and disclose records of the source of copyright works
- The TDM exception will not be applicable if relevant licensing schemes are available
- An “opt out” option is provided for copyright owners

AVIA would suggest that the above guardrails are **insufficient** to give protection to copyright owners. Furthermore, we are doubtful about how these restrictive conditions can be implemented in a manner that can truly safeguard the rights of content owners. Specifically, IPD proposes an opt out system to allow content owners to reserve their rights to content they do not want to be permit training on. However, such an opt out system is antithetical to copyright, which recognizes the exclusive right of an author to authorize the use of their work.

Additionally, for any rights reservation system to be effective, right holders need to be able to easily to exclude their works from training in an effective and non-burdensome manner.

IPD’s statement to LegCo that “the TDM exception will not be applicable if the copyright owners have expressly reserved their rights, such as indicating their choice to reserve their rights through machine readable means in the case of content made publicly available online.” However, we would highlight that there is at present no known effective means of delivering the ability for a rights holder to “reserve their rights through machine readable means”. Robots.txt is ineffective. At best it prevents crawling of a site that is controlled by the rights holder but does nothing to prevent the same content from being scraped from other unauthorized (or even in some cases authorized) sources, whether YouTube or the PirateBay. It can also result in sites choosing to opt-out from AI-training also removing their sites from search results when dealing with retrieval-augmented generation.

What is needed — and does not currently exist — is a mechanism to give effect to the choice of a right holder to exclude their work from training on a per-work basis, through the use of persistent metadata, or through reference data similar to what is done in the case of ContentID and similar technologies used to enable pre-upload filtering on UGC services.

Furthermore, for this to be an effective safeguard, regulators would need to confirm that an opt-out system actually works to give effect to the choices of rights holders before implementing a TDM exception that depends on an opt-out as the means of protecting authors.

Should a TDM exception be introduced for training AI systems/models and services that are offered to the general public or external commercial customers, AI developers should be required to maintain reasonable records of training data, including as to the use of copyright protected works, and make those records easily and readily available for review, to provide

transparency to rightsholders and users. In this regard, a requirement for AI Platforms to introduce measures to ensure developers are transparent about the works their models are trained on could be introduced to help underpin existing copyright law. Reasonable transparency measures would help achieve the intended balance of objectives by facilitating innovation through accountability and building trust and public confidence, and provide regulators with the ability to verify compliance with any restrictions on training, and providing right holders with information to help them enforce their rights.

Additionally, AI developers should be required to maintain reasonable records of training data, including copyright protected works in furtherance of a number of important policy objectives. To help enable copyright enforcement, transparency on the following information relating to training data would be necessary:

- Online resources used, such as websites, online service, platforms, etc., including different specific content pages.
- Third-party databases/data sources/datasets used.
- Type of data and/or type of content concerned (audio, video, image, etc., including copyright-protected content) in a way that is detailed enough to capture content that includes other content (e.g., for images of consumer products, what characters are present; and, for social media, what trailers or pirated content were included).
- Concrete timeframe(s) during which the specific data/content concerned was collected.

This information should be provided in a dynamic manner, with updates made in real time when material changes occur.

Finally, AVIA proposes that it should not be expected that all content should be made available for training of AI systems, **nor should one assume that it needs to be in order to develop an AI platform**. Some rightsholders of high-value IP might reasonably be expected to withhold their works from training. Some rights holders will be interested in licensing revenue, while others will prioritize the control over their IP that copyright allows in this space. That is what copyright law envisions and that does not threaten the ability of AI platforms to innovate. Hong Kong's proposed exemption to allow commercial use is out of line with existing international jurisdictions who have already introduced TDM exceptions but have, importantly, restricted it to non-commercial use. We would urge Hong Kong to reconsider this forward leaning approach which is unnecessarily ahead of international trends.

## **Conclusion**

Whilst we recognise and support the Hong Kong government's decision to encourage AI company development, we are concerned that legislating a TDM exception would not be beneficial to the creative industry. AVIA believes that the existing Copyright Ordinance is sufficient to ensure copyright protection whilst allowing AI companies to negotiate access, on a

case-by-case basis with rightsholders for TDM purposes. Introducing a specific TDM exception, with unenforceable guardrails offers limited to no protection for the creative industry.