BEFORE THE
COMMERCE AND ECONOMIC DEVELOPMENT BUREAU AND
INTELLECTUAL PROPERTY DEPARTMENT OF HONG KONG

Public Consultation on Copyright and Artificial Intelligence (2024)	**06 September 2024**

The Copyright Alliance appreciates the opportunity to submit comments to the Commerce and Economic Development Bureau (CEDB) and Intellectual Property Department (IPD) of Hong Kong in response to the Public Consultation on Copyright and Artificial Intelligence (2024) ("Consultation"), specifically concerning Chapter 4 of the Consultation and the proposal for a tailored exception for text and data mining ("Proposed TDM Exception").[1] Specifically, the CEDB and IPD asks:

> "whether it is justifiable to introduce into the CO a new and specific TDM exception for the purposes of computational analysis and processing of text, images, data and/or other types of information, which shall cover (a) conventional text and data mining; and (b) computational data analysis and processing for enhancing the performance of a computer program such as the development, training and enhancement of AI models."[2]

---

[1] GOV'T HONG KONG SPECIAL ADMIN. REGION, COMMERCE AND ECON. DEV. BUREAU & INTELL. PROP. DEP'T, COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PUBLIC CONSULTATION PAPER (2024), https://www.ipd.gov.hk/filemanager/ipd/en/share/consultation-papers/Eng-Copyright-and-AI-Consultation-Paper-20240708.pdf (last visited Sept. 6, 2024).

[2] *Id*. at 32.

The Copyright Alliance is a non-profit, non-partisan public interest and educational organization representing the copyright interests of over 2 million individual creators and over 15,000 organizations in the United States, across the spectrum of copyright disciplines.[3] We are dedicated to advocating policies that promote and preserve the value of copyright, and to protecting the rights of creators and innovators who rely on copyright law to protect their creativity, efforts, and investments in the creation and distribution of copyrighted works for the public to enjoy.

The Copyright Alliance and our members support the responsible, ethical, and respectful development and use of AI technologies. The continuing development of AI systems brings many opportunities. Many in the creative industries are already using or plan to use AI-based technologies to assist in the creation of a wide range of works that benefit society. In fact, some—like the motion picture, video game, and music industries—have been using AI-based assistive tools for many years. Others—like many independent illustrators and authors—have just begun exploring how to incorporate AI tools into their work process. For example, creators can use AI tools to assist with ideation, artists can use AI image generators to combine elements with original artwork, and digital media licensors and technology companies developed their own image generative AI ("GAI") tools.[4]

The development and use of AI systems also bring many challenges, especially related to copyright. Copyright laws must not be cast aside in favor of new policies obligating creators to effectively subsidize AI technologies under the misguided belief that doing so is necessary to incentivize AI technologies. It is also essential that those using copyrighted works to develop AI systems not devalue the rights and interests of creators and copyright owners and undermine copyright protections. This is especially true where there is no evidence of market failure or problems warranting measures such as the current proposal. Legal change should not precede commercial reality.

---

[3] A full list of Copyright Alliance organizational members is available online. *See Who We Represent*, COPYRIGHT ALLIANCE, https://copyrightalliance.org/about/who-we-represent/ (last visited Sept. 5, 2024).

[4] *See AI Generator*, GETTY IMAGES, https://www.gettyimages.com/ai/generation/about (last visited Sept. 5, 2024); *Adobe Firefly Overview*, https://helpx.adobe.com/firefly/using/firefly-overview.html (last visited Sept. 5, 2024).

Copyright law empowers independent to large-scale creators and rights holders to create inspiring, innovative, and pioneering works. These works not only benefit the public, but they can also help drive and benefit AI development and use, positively contributing to the economy and employment. A 2023 study on the economic impact of intellectual property in Hong Kong by the IPD reported that the copyright-intensive sector accounted for HK$132.7 billion or 4.9% of Hong Kong's GDP[5] and 7.2% of total employment in Hong Kong.[6] It is vital for any proposed AI-related copyright policies to take into account the effect such policies may have on copyright's impact to the economy and job creation.

Like many other countries, including the United States, the Hong Kong government is considering whether legal or policy changes are necessary or appropriate to foster AI innovation. More specifically, in the Consultation, the CEDB and IPD are considering whether changes to Hong Kong's Copyright Ordinance ("CO") are appropriate. While fostering AI innovation is a worthy goal, obtaining that goal should not be accomplished at the expense of Hong Kong's and other countries' creative communities nor should it undermine the intellectual property laws that support them. It is vital that any revisions to Hong Kong law respect intellectual property—and in particular copyright. Unfortunately, as explained in more detail below, the Proposed TDM Exception fails to meet this goal and seems to disregard the value and virtues of strong and effective copyright laws. We, therefore, oppose the Proposed TDM Exception and urge CEDB and IPD to not adopt it.

Creators and rights holders around the globe are deeply concerned about the negative impacts the Proposed TDM Exception would inflict on creativity and the creative community. The Proposed TDM Exception undermines the fundamental rights of creators and copyright owners, potentially violates international IP obligations, and runs counter to other countries' approaches to AI, including the United States. Any changes to law and policy must be justified by the existence of

---

[5] *See* GOV'T HONG KONG SPECIAL ADMIN. REGION, INTELL. PROP. DEP'T, STUDY ON CONTRIBUTION OF INTELLECTUAL PROPERTY-INTENSIVE INDUSTRIES TO HONG KONG ECONOMY 20 (2023), https://www.ip.gov.hk/filemanager/ip/en/content_150/Study-on-IP-Intensive-Industries-to-HK-Economy-e.pdf.

[6] *Id.* at 26.

a problem to be solved. Right now, no evidence exists to support the existence of a problem. As we detail further below, the text and data mining ("TDM") licensing markets for copyrighted works have existed for a long time and continue to flourish with the rise of GAI. Adopting the Proposed TDM Exception would adversely impact these developing markets and undermine fundamental rights of creators and copyright owners around the globe. GAI developers rely on rich, large and quality datasets for AI training that may be commercialized. Therefore, it is vital to ensure that human creators (i) continue to have sufficient protections for their works used in such training systems, (ii) are remunerated for their works that are used for AI training, and (iii) remain incentivized to continue creating, which in turn will lead to greater volume and diversity of works that can be licensed for use as GAI training materials.

We address the following questions from the Consultation in these comments.

> *What further justifications and information can be adduced to support (or roll back) the idea of introducing the Proposed TDM Exception into the CO with a view to incentivising the use and development of AI technology and pursuing overall benefits? Is copyright licensing commonly available for TDM activities? If so, in respect of which fields/industries do these licensing schemes accommodate? Do you find the licensing solution effective?*

Copyrighted Works Drive Development of AI Technologies

We urge CEDB and IPD to withdraw the Proposed TDM Exception. We cannot overemphasize the devastating harms that the exception would have on creators and rights holders and the disruption it would cause to the TDM licensing marketplace. A broad TDM exception like the current proposal would severely undermine both existing and developing licensing markets for the use of copyrighted works for AI training purposes.

Copyright law is not a barrier for the use and development of GAI technologies. Quite the opposite—copyright law enhances and fuels the development of GAI technologies. The reason that the creative community, from independent to large-scale creators and rights holders, is able

to create high-quality works is because of strong copyright laws. And it is these high-quality works (that often require significant investment by the creator) that are ideal to train AI machines to generate high-quality output, including through techniques such as TDM.

Many U.S. creators and rights holders, particularly publishers and image/media licensors, already license their copyrighted works for commercial AI uses and many of those that do not are on the cusp of doing so. Just a few public examples of licensing solutions, initiatives, and agreements for AI use of copyrighted works include those launched from or created by [Created by Humans], [Dataset Providers Alliance], [Copyright Clearance Center], [Elsevier], [Getty Images], [Shutterstock], [Jstor], [Sage Journals], [Rightsify], [Universal Music Group], and other major media publishers including [News Corp], [Associated Press], [The Atlantic], [Vox Media], [Dotdash Meredith], [Financial Times], [Fortune, Time, Entrepreneur, The Texas Tribune, and WordPress.com]. Through such licensing deals, copyright owners not only provide high quality copyrighted works for better AI training and development, but also make these copyrighted works useful for various AI-based use cases, including scientific research, through semantic enrichment, metadata tagging, content normalization, and data cleanup.

As copyrighted works are vital to the development of these technologies, Hong Kong should require AI developers and deployers to first obtain appropriate licenses and authorizations in connection with any works they desire to ingest for purposes of AI development, and before deploying any AI system that was developed using such works. Those policies should also require AI developers and deployers to maintain records of copyrighted works used to train an AI system. In addition, Hong Kong should promote and invest in proper education on the ethical development and deployment of AI, including education on copyright and the importance of seeking authorization before ingesting copyrighted works.

GAI systems trained on copyrighted works are often used to generate AI outputs that displace the market for the very copyrighted works on which they are trained. Since GAI can generate outputs that can displace ingested works, the Proposed TDM Exception *only* benefits AI developers since it obligates human creators and rights holders—who are rarely consulted for

approval or compensated for their works—to subsidize AI developers. The proposal therefore would greatly devalue the copyrights of all creators and rights holders.

Additionally, creators and rights holders are concerned that the Proposed TDM Exception will increase the risk of piracy and loss of proprietary information and methods, especially where the ingested copyrighted works contain technical protection measures or data/metadata relating to individual customers or other sensitive information and methods.[7] Unrestricted TDM use, as proposed by CEDB and IPD, increases the risk of circumvention and mishandling of data and technologies subsisting in the copyrighted works.

Proposed Conditions to Narrow the Proposed TDM Exceptions Are Inadequate in Addressing Rights Holders' Concerns

CEDB and IPD offer several possible conditions to place on the Proposed TDM Exception as "viable option[s] for safeguards" for rights holders.[8] We address each condition in turn, noting why none of them is a viable option as a safeguard for rights holders:

Lawful Access: The Consultation poses a requirement that TDM users must have "lawful access" to the copyrighted works. This does not alleviate creators' and rights holders' concerns. For example, a "lawful access" requirement would not address instances when authorized or pirated copies of copyrighted works are uploaded to publicly accessible websites that are then scraped by an AI company and used for AI training purposes. Simply making a work available online should not allow an AI developer or anyone else to copy and make use of that work as they fit. To conclude otherwise would obliterate well-established tenets of copyright law. In response to the lawful access condition, creators and rights holders may be forced to restrict access (such as by placing copyrighted works behind a paywall), considerably reducing the availability of high-

---

[7] For example, training of images featuring people invariably involves the processing of personal data in the form of biometric and personal information that trigger privacy and other related concerns. By properly licensing the images, AI developers can address these concerns through model release agreements already secured for those images by creators and rights holders. This is not possible if the images were to be taken and used under a TDM exception because an AI developer would need to separately secure separate model release agreements for every individual in an image.

[8] GOV'T HONG KONG SPECIAL ADMIN. REGION, *supra* note 1 at 37.

quality content to the public. Restricting access to works that are presently widely accessible is not in anyone's interest.

Opt-Out: A second proposed condition states that TDM activities under the Proposed TDM Exception would not apply if a TDM license is available for the work or the copyright owner has expressly reserved their rights (i.e., an opt-out requirement). This opt-out condition subverts the foundational rule of copyright law that a copyright owner has the right to choose whether to authorize others to use their work. Copyright owners should not be required to take affirmative action to prevent others from using their works. As an opt-in regime, copyright law appropriately places the onus on copyright users, such as AI companies, who are in a far better position to determine and discern which works they want to use and to secure authorizations to do so. The marketplace should continue to properly value and incentivize creativity, and AI policy should not interfere with the right or ability of copyright owners to license, or choose not to license, their works for AI purposes. A copyright owner should be free to decide whether they want to license their work to an AI company (or anyone else).

Additionally, requiring rights holders to opt out of AI training or provide a TDM license would effectively impose a copyright formality on rights holders. This requirement would violate Article 5 of the Berne Convention, which states that copyrights are to be enjoyed and exercised by authors without being subject to any formality.[9]

Though a few AI companies have offered opt-out options, many do not. Even those opt-out schemes that do exist are ineffective for various reasons. For example: (i) the copyrighted works might already have been copied and used for training at the time of opt-out; and (ii) despite opting out, copies of the copyrighted works may still be included in the datasets through other means, such as when copies are scraped from other sources such as a licensee of the copyright owner or from a third-party platform where a copy has been posted. The practical effects of opt-out, particularly with regards to works already used to train AI, are also negligible given that it is challenging to remove entire works at scale from an AI model, particularly for an AI model that

---

[9] Berne Convention for the Protection of Literary and Artistic Works, art. 5(2), Sept. 9, 1886, as revised at Stockholm July 14, 1967, 828 U.N.T.S. 221, 231-32.

has already been trained on works where the copyright owner has opted out. While the Consultation mentions that existing technical solutions may assist with opt-out, these tools typically have significant limitations because they are only effective to the extent opt-out is recognized and respected, and because these tools are often not designed to be targeted to address scraping for GAI ingestion.[10] Moreover, copies of works that are available on pirate sites are outside the copyright owner's control. Allowing broad-scale web scraping means the work will end up in a training dataset even if the copyright owner has opted out. For all these reasons, there is currently a high level of uncertainty in the European Union over what constitutes effective opt-out, and as time passes this uncertainty is being exploited by AI developers who continue to train on scraped content despite legitimate efforts form copyright owners to opt out.

Output Restrictions: The third proposed condition states that the Proposed TDM Exception could be conditioned on imposing restrictions on further communication, distribution, and dealing of the copy made under the exception. Such a condition would be completely ineffective where the entity relying on the Proposed TDM Exception is also the one undertaking the AI training. Once the training is complete, it is not necessary for a copy of the copyrighted work to be further communicated, distributed, or dealt with in order to commercialize the resulting AI model. While this condition might in some very limited circumstances hypothetically address harms arising from infringing AI output, implementation of this condition would be near-impossible to monitor or enforce. In any event, the condition fails to address the harmful effects that the Proposed TDM Exception will have on existing and developing TDM licensing markets.

---

[10] Robots.txt protocol is one example. While robots.txt does alert scraping tools not to ingest the associated copyrighted work, it has significant limitations because it is only effective to the extent it is recognized and respected, and it was not designed to be targeted to scraping for generative AI ingestion. Robots.txt may also prevent a search engine from indexing the work. A copyright owner may want their work to be scraped for search engine purposes—so they can be found on the internet—but not for AI ingestion. Even if robots.txt is used, it does not attach to the copyrighted work itself but will operate at the URL or website level.

U.S. and International Approaches to TDM Exceptions

For many years now, lawmakers and policymakers in many countries, including the United States, have been carefully examining the intersection of copyright law and AI.[11] Most countries, including the United States, have not enacted any TDM exceptions or new exceptions to copyright law for AI purposes. And for good reason: AI licensing markets in the United States are robust and, absent contrary evidence, uninformed or ill-timed legislation may cause significant harms to such markets.

Additionally, arguments that unauthorized uses of copyrighted works (including TDM) may be permitted under U.S. copyright law, particularly on fair use grounds, are overgeneralizations of U.S. law. U.S. copyright law requires courts to engage in a fact-specific analysis using an established four-factor test and forbids categorically excusing otherwise infringing uses of copyrighted works. As such, any AI use, must be carefully analyzed in context. Because fair use analysis is *always* a fact-intensive test tailored to a specific situation it is an unreliable benchmark to justify sweeping new legal changes or the creation of new copyright exceptions. There can be no broad characterizations of what AI uses of copyrighted works are allowed in the United States because our laws simply do not provide for *any* such broad exceptions.

The Consultation notes that "a number of overseas jurisdictions have updated their copyright laws to provide for a specific copyright exception, using a general label of text and data mining exception."[12] However, in fact, only a few jurisdictions have enacted express TDM exceptions within their copyright laws, including Singapore, Japan, and the EU. Meanwhile, the UK has a limited exception that excuses TDM of lawfully accessed copyrighted works for the sole purpose of research for a non-commercial purpose. More recently, a proposed expansion of the UK TDM

---

[11] U.S. federal agencies that have examined the issues include the National Institute of Standards and Technology (NIST) and the U.S. Patent and Trademark Office (USPTO). *See* Study to Advance a More Productive Tech Economy, 86 Fed. Reg. 66287 (Nov. 22, 2021), https://www.regulations.gov/document/NIST-2021-0007-0001; Request for Comments on Intellectual Property Protection for Artificial Intelligence Innovation, 84 Fed. Reg. 58141 (Oct. 30, 2019), https://www.regulations.gov/document/PTO-C-2019-0038-0001.

[12] GOV'T HONG KONG SPECIAL ADMIN. REGION, *supra* note 1 at 31-32.

exception failed to gain any traction and was pulled by the UK government in 2023 over grave concerns that it would significantly devalue copyright and severely harm rights holders.[13]

Other countries have each considered whether TDM exceptions should be introduced into their copyright laws, and in each case, have either declined to take action, postponed making a decision as premature, or otherwise not taken action. These countries include South Korea, New Zealand, Australia, and Canada.

Indeed, evidence demonstrates the existence of strong and vibrant AI licensing markets, soundly refuting the need for legal changes that favor AI developers at the expense of rights holders and creators. The desire to be the leaders of technological innovations cannot compromise the foundations that allow for such innovations to occur in the first place. As previously stated, without strong copyright laws that incentivize and protect the creation and dissemination of copyrighted works, there cannot be trustworthy, reliable, and ethical AI technologies.

The Proposed TDM Exception Fails to Comply with Berne Convention

The Proposed TDM Exception also clearly fails to pass the three-step test of the Berne Convention and the World Trade Organization Agreement on Trade-Related Aspects of Intellectual Property (TRIPS Agreement), to which Hong Kong is a signatory.[14] Enshrined in Article 9(2) of the Berne Convention and Article 13 of the TRIPS Agreement, but traditionally broken down into three components, the test provides that member countries may permit the

---

[13] On February 1, George Freeman—Minister for Science, Research, and Innovation—announced that the United Kingdom government will take a step back from its original proposal put forth in the summer of 2022 for a broad exception for the text-and-data mining of copyrighted works for any purpose in the UK's copyright laws. The Minister noted that the government will further consult stakeholders in the coming months to "ensure that we do not rush precipitately into a knee-jerk move that is wrong." *See* HC Deb (1 Feb. 2023) (727) cols. 152-68WH (UK), https://hansard.parliament.uk/commons/2023-02-01/debates/7CD1D4F9-7805-4CF0-9698-E28ECEFB7177/ArtificialIntelligenceIntellectualPropertyRights.

[14] Wu Jianmin, *Berne Notification No. 186, Berne Convention for the Protection of Literary and Artistic Works, Application of the Berne Convention, with effect from July 1, 1997, to the Hong Kong Special Administrative Region*, WIPO (July 7, 1997), https://www.wipo.int/treaties/en/notifications/berne/treaty_berne_186.html [https://web.archive.org/web/20240412220832/https://www.wipo.int/treaties/en/notifications/berne/treaty_berne_186.html].

reproduction of copyrighted works in (1) "certain special cases," (2) "provided that such reproduction does not conflict with a normal exploitation of the work," and (3) "does not unreasonably prejudice the legitimate interests of the author."[15]

The Proposed TDM Exception fails under the very first step of the test because it is not a "special case." The exception is extremely broad in its scope as it excuses *all* unauthorized TDM uses conducted for *any* purpose—the exception is not limited or narrowed further in any way. The CEDB and IPD suggest possible conditions to limit the Proposed TDM Exception. But as explained above, the suggested safeguards, while well-meaning, ultimately fail to address key concerns with the Proposed TDM Exception, resulting in the exception being still much too broad.

The Proposed TDM Exception also fails the second step of the test because it would directly undermine the normal exploitation of copyrighted works for TDM uses. As noted above and throughout this submission, rights holders and creators are currently licensing content to AI developers, and the creative community continues to explore new ways that they can license copyrighted content for AI training purposes. A proposal undermining that licensing market would directly conflict with the existing and normal exploitation for the works and prejudice authors' legitimate interests in their works.

Moreover, authors' legitimate interests are severely prejudiced where the outputs of an AI model trained in reliance on a TDM exception act as a substitute for and compete against the copyrighted work that has been copied. An output does not need to be identical to be a plausible substitute. The quality of generative AI outputs already available on the marketplace can make it hard to distinguish between human-created works and AI generated output and lead to a total displacement of human-created works. The longer-term displacement potential of the Proposed TDM Exception is obvious and not in the interests of the public. The proposed conditions regarding the existence of a TDM license or opt-out, at best, only partially helps alleviate some of these concerns. But ultimately,

---

[15] Berne Convention for the Protection of Literary and Artistic Works, art. 9(2), Sept. 9, 1886, as revised at Stockholm July 14, 1967, 828 U.N.T.S. 221, 239.

the Proposed TDM Exception would still directly undermine the legitimate interests of authors to enjoy and exploit their copyrighted works.

Overall, as discussed earlier, the Proposed TDM Exception would stunt growth in the AI licensing market and prejudice rights holders' fundamental rights and ability to license and be compensated for their protected works. Undermining such a fundamental pillar of copyright law in favor of a new technology is precisely the kind of scenario the three-step test protects against, and that Berne Convention and TRIPS Agreement member countries have agreed to avoid.

For the reasons cited above, we urge the CEDB and IPD to withdraw the Proposed TDM Exception.

Respectfully Submitted,

Keith Kupferschmid
CEO
Copyright Alliance
1331 F Street, NW, Suite 950
Washington, D.C., 20004