

Electronic submission via: AI_consultation@cedb.gov.hk

September 6, 2024

Division 3
Commerce and Economic Development Bureau
23rd Floor, West Wing
Central Government Offices
2 Tim Mei Avenue
Tamar, Hong Kong

Subject: Copyright and Artificial Intelligence: Public Consultation

The Association of American Publishers (AAP), a not-for-profit organization, represents the leading book, journal, and education publishers in the United States on matters of law and policy, advocating for outcomes that incentivize the publication of creative expression, professional content, and learning solutions. AAP's members range from major commercial book and journal publishers to small, non-profit, university, and scholarly presses, as well as leading publishers of educational materials and digital learning platforms—publishers that curate and bring to market works that educate, entertain, and inform.

This submission responds to the questions posed in the Public Consultation Paper, but ultimately rests on two key principles:

1. The wholesale reproduction of copyrighted works for purposes of training and developing AI systems, absent the copyright owner's authorization, is infringement.
2. Any framework intended to promote AI development must not diminish the copyrights of the authors and publishers whose works are essential to a flourishing and well-informed society.

Copyright is the engine of self-expression, responsible for incentivizing invaluable, creative expression and information that is inherently transformational to both people and society and indisputably essential to a well-functioning society. Copyright protection also fuels markets. Human authorship is more important than ever in the age of artificial intelligence.

A vibrant creative sector is necessary for AI to reach its full potential. It is not and should not be a copyright owner's burden to subsidize the development of generative AI technologies, and government should certainly have no role in bestowing legal or commercial advantages to AI companies at the expense of authors, publishers, and other creators—without whose works AI models would not be possible.

The companies that benefit from the commercialization of this technology should be required not only to compensate rights holders for their past ingestion of copyrighted works to train generative AI systems but also for their ongoing and future use of protected works to train new generative AI systems or fine-tune their existing products—including through text-and-data mining activities.

AAP appreciates the opportunity to submit the below responses to questions posed in the Public Consultation Paper. Generative AI systems may be poised to accomplish important advances, but their development and the economic dominance of the companies that develop these technologies should not be had at the expense of the creative sectors on whose works the technologies' underlying large language models have been trained. The Hong Kong government should not sanction the ingestion of valuable intellectual property unconstrained.

Chapter 2: Copyright Protection of AI-generated Works

1. Do you agree that the existing Copyright Ordinance offers adequate protection to AI-generated works, thereby encouraging creativity and investment, as well as the usage, development, and investment in AI technology?

There is merit in considering whether and how AI-generated outputs that are *elements or aspects of human created works* may be included in the copyright protection accorded to works created through human authorship. However, it is our view that outputs entirely generated by a generative AI algorithm or technology—that is, one entirely devoid of human authorship—is not and should not be the subject of copyright protection. The purpose of copyright, as the consultation paper rightly notes, is to provide “economic incentives for the creation¹” of works *by humans*. Legal protection for AI-generated material would not serve the purpose of copyright law. Copyright affords authors and publishers legal protections for the intellectual property—the works—they produce, which encourages the creation, commercialization, and dissemination of new works to the public. This legal protection assures the artist, the author, the rights holder that their investment, whether intellectual or otherwise, in the creation and dissemination of the creative work will see a return that allows them to continue to invest in the process of creation, in the distribution of such works to the public, thereby enriching culture and social discourse—through “access to diverse and creative works.”²

Generative AI tools or systems require no such incentives, and the output of these systems, on their own, lack the type of human creative expression on which copyright is premised. As such, material or output solely generated by a generative AI system, completely absent human involvement accomplishes no policy imperative to grant copyright protection to the output of an AI algorithm. Indeed, copyright protection appears irrelevant to the developers of generative AI systems, which companies have invested and continue to invest in further developing this technology. Their investment has not been and is not tied to the grant of legal protections such as copyright to the output of the generative AI system they may have developed.

It is also the case that these generative AI tools are capable of producing outputs at a rate no human creator can match—not only flooding the market with outputs that may unfairly compete with the works created by the human author and artist, but which may also devalue the content created through human authorship.

On the question of whether the Copyright Ordinance is sufficient to address the question of protection for AI-generated outputs that may be included in human created content, we agree with

¹ Copyright and Artificial Intelligence: Public Consultation Paper, p. 7 (“Consultation Paper”).

² *Id.*, at p. 8.

the government’s view that it is not necessary to propose any substantive legislative amendments to address this question—at this time. What is critical is that human authorship remain paramount when considering whether copyright protection should be accorded. However, what remains the difficult question is what should constitute “*sufficient* human authorship” as to accord copyright protection to AI-generated outputs embedded as elements of a copyrightable work.

- 2. Have you relied on CGWs provisions of the CO in the course of claiming copyright protection for AI-generated works? If so, in what circumstances, how and to what extent has human authorship featured in these works? Have you experienced any challenges or disputes during the process?**

Not answered.

- 3. Do you agree that the contractual arrangements in the market provide a practical solution for addressing copyright issues concerning AI-generated works?**

Questions regarding ownership of AI-generated outputs can be more clearly defined in the licensing arrangements between the parties. Perhaps analogous to the manner in which a company acquires copyright ownership over the works created by its employees, under a work for hire arrangement, an entity employing a generative AI system to accomplish defined tasks may similarly obtain ownership of the AI system’s output if so agreed under the licensing agreement between the parties.

Licensing agreements or contractual arrangements remain the best tool for facilitating AI development while also protecting the rights of authors, publishers, and other copyright owners and licensees, but can also provide flexibility, while affording the parties greater stability and certainty with respect to their rights and obligations on a number of copyright-related concerns, such as the scope of use, authorship, and ownership of the training data sets and the AI-generated outputs.

Chapter 3: Copyright Infringement Liability for AI-generated Works

As the consultation document notes, the focus of this chapter is potential infringement liability for the creation and use of AI-generated works,³ or rather, outputs. However, what the consultation document does not seem to address is the infringing activity that has occurred, and likely continues to occur, when a generative AI developer creates and uses training datasets from the copyrighted works of creators and authors but for which neither permission or authorization is obtained, nor compensation paid. Per the consultation document,⁴ “(c)opyright in a work is infringed by a person who, without license of the copyright owner, does or authorizes another to do any of the acts restricted by copyright in relation to the whole or a substantial part of the work, unless the act in question is permitted under any statutory copyright exception(s).” To our mind, it appears that it is the view of the government that the wholesale reproduction and ingestion of copyrighted works for training generative AI, without permission or compensation to the copyright owner, is infringing conduct—a view we heartily endorse.

³ Id, at p. 22.

⁴ Id.

1. ***Do you agree that the existing law is broad and general enough for addressing the liability issues on copyright infringement arising from AI-generated works based on the individual circumstances?***

The question of liability for copyright infringement arising from AI-generated works appear to be readily addressed under existing law. Where AI-generated material is suspected to infringe a copyrighted work, the end user of the system — for instance, the party that crafted and refined the descriptive text prompts or directions that instruct the AI system to generate the output should be directly liable for infringement. Both the developer of the Gen AI model and the developer of the system incorporating that model into another product or service, depending on the facts, may be held either directly liable or secondarily liable under the relevant theory of secondary liability. As the government concludes, “(t)his is no different from the long-standing fact finding approach we have adopted for determining copyright infringement involving non-AI generated works.”⁵

2. ***Have you experienced any difficulties or obstacles in pursuing or defending legal claims on copyright infringement arising from AI-generated works? If so, what are such difficulties or obstacles?***

Not answered (as AAP has not brought an action in Hong Kong to allow it to address this question).

3. ***Do you agree that the availability of contractual terms between AI system owners and end-users for governing AI-generated works also offers a concrete and practical basis for resolving disputes over copyright infringements in relation to these works? If not, could you share your own experience?***

Contractual arrangements may indeed be the better vehicle through which to define the relationships between the generative AI developer and the entities that employ the generative AI system within their own business processes. It is our view that if an output generated by an AI process infringes a copyrighted work, the infringement can be ascribed to either the human programmer or the human user of the AI system or algorithm, i.e., the human actor who ultimately either designed (i.e., programmed the desired output or task) or engaged the AI model to produce the infringing output. Questions may arise as to whether the human actor is directly liable or merely contributed to the infringing conduct, in which case, the same analysis and assignment of liability under existing law appears sufficient to address this question. At its core, it is not a question of whether a party is exercising control over the generative AI technology, but instead, which party set the infringing conduct in motion and should therefore be held accountable.

Chapter 4: Possible Introduction of Specific Copyright Exception

1. ***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?***

The Copyright Ordinance appears sufficiently robust in its ability to respond to new technologies and evolving user needs—a point the consultation paper itself makes repeatedly.

⁵ Id, at p. 23.

Indeed, the consultation papers offers no evidence of a market failure that would justify upending the current framework, which remains capable of meeting the needs of users (intending to and engaging in TDM activities) while continuing to protect the creative sector — upon whose creative works depends the development and training of generative AI systems. There appears to be no reason the current copyright framework cannot accommodate and support the continued development of AI. While the consultation paper posits that “(i)ntroduction of the Proposed TDM Exception is conducive to attracting more I&T enterprises and talents to invest and engage in AI industries in HK,” such an assertion is largely speculative. In the jurisdictions in which an exception already exists, there have been no publicized reports that the introduction of such an exception has indeed brought in the anticipated influx of investment in AI technologies.

While a few jurisdictions have adopted a copyright exception for text-and-data mining, on the premise that it is necessary to AI research and development, AAP does not believe any exceptions to permit the unlicensed use of copyrighted works to develop AI technologies are necessary or desirable. We note that exceptions for TDM activity (where they exist) were adopted before the rise of generative AI, and the original aim was to facilitate computational analysis of large amounts of text as may be embodied in scholarly and scientific articles for research purposes. TDM uses do not equate with the use of copyrighted works to train generative AI models. If the objective is to facilitate TDM activity that may be helpful to the further development of an AI industry, licensing solutions remain the better tool for facilitating such activity, and in turn, development of AI technologies. Through licensing arrangements, AI systems developers can legitimately access the copyrighted works necessary to training trustworthy and reliable AI models while authors, publishers, and other copyright holders and licensees are appropriately compensated for the use of their works.

A purported justification for introducing the Proposed TDM Exception is the supposed inconvenience of respecting the rights of creators or compensating them for the use of their works, of having to “obtain consent from different copyright owners.”⁶ However, it should not be the burden of publishers and other rights holders to subsidize the activities of entities that intend to profit from the findings yielded by their “mining” of databases comprised of copyrighted works whose creation, publication, and dissemination is made possible by the investments of publishers and other rights holders. Mere isolated instances of difficulties are not indicative of a widespread lack of access that necessitates the introduction of a TDM exception.

We note that the costs that might be associated with obtaining permission and compensating rights holders for the use of their works in a TDM activity are not only relatively minor compared to the overall expenditures of Gen AI developers, they are but among a long list of costs that influence where a company chooses to locate its business. Among many, other costs include availability of a robust talent pool, quality of the technological infrastructure, proximity to universities with strong programs in relevant fields, access to private equity and venture capital, proximity to customer markets, local languages, servers, network infrastructure, real estate, research and development, energy, legal, finance, and human resources.

2. *How would the Proposed TDM Exception overcome the obstacles/limitations you have experienced in conducting TDM activities and facilitate the development of your business and industry?*

⁶ Id. at p. 35.

Not answered.

3. 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?

Licensing is indeed already available for TDM activities. While there are as yet no comprehensive economic studies on the value of the global text-and-data mining (TDM) licensing market (which we note is not equivalent to licensing for generative AI purposes),⁷ it should be noted that scholarly and professional publishers already license their journal databases for such TDM activities,⁸ and publishers of consumer/trade books and publishers of educational materials are also engaged in licensing discussions.

The market exists for commercial TDM of published works and continues to grow. With the market continuing to evolve, it would be premature to intervene in its evolution — such as by adopting an unnecessary and broad exception, given the scant evidence to justify its adoption. Publishers and other rights holders continue to experiment with novel licensing options — based on the exclusive rights afforded by copyright law — to meet evolving user demands. As such, AAP strongly cautions against the introduction of the Proposed TDM Exception.

We note that licensing arrangements remain the best mode through which TDM activities, and in turn, AI development, may be facilitated, while also protecting the intellectual property rights of the creators and curators of copyrighted works that may be used to train and develop AI technologies. Licensing can better provide the flexibility desired by the parties, with novel licensing arrangements designed to address the bespoke needs of licensees and licensors.

4. What conditions do you think the Proposed TDM Exception should be accompanied with, for the objective of striking a proper balance between the legitimate interests of copyright owners and copyright users, and serving the best interest of Hong Kong? Are there any practical difficulties in complying with the conditions?

We do not find a Proposed TDM Exception to be necessary. Mere inconvenience is an insufficient justification for weakening the exclusive rights accorded to content creators and rights holders, particularly where licensing mechanisms are available to facilitate TDM activities.

However, if the government were inclined to adopt such an exception, of the regimes discussed in the consultation paper, we note that the EU and UK frameworks are the least objectionable. Both frameworks require lawful access to the work, differentiate TDM activities for

⁷Text Mining Market Trends: Growth & Analysis, 2020 – 2027 (Oct. 2023), <https://www.reportsanddata.com/report-detail/text-mining-market>.

⁸See e.g., *Text & Data Mining: Harness the Power of Big Data and Analytics*, ACS PUBLICATIONS (Nov. 23, 2022), <https://solutions.acs.org/solutions/text-and-data-mining/>; Elsevier Text and Data Mining (TDM) License, ELSEVIER, <https://beta.elsevier.com/about/policies-and-standards/text-and-data-mining/license?trial=true>; Sage Journals Text and Data Mining License Agreement, SAGE JOURNALS, <https://journals.sagepub.com/page/policies/text-and-data-mining-license>; *Text Data and Mining*, TAYLOR & FRANCIS, [Text and Data Mining - Taylor & Francis \(taylorandfrancis.com\)](https://www.taylorandfrancis.com/text-and-data-mining/); *Text Data and Mining*, WILEY, <https://onlinelibrary.wiley.com/library-info/resources/text-and-datamining>.

commercial and non-commercial purposes (specifically, for scientific research), and in the EU framework, the exception for commercial TDM activity applies unless the rights holder prohibits the use of their copyrighted works by expressly reserving their rights “in an appropriate manner, such as machine-readable means in the case of content made publicly available online.”⁹

We encourage CEDB to avoid the formulation of the TDM exceptions as adopted in the copyright laws of Japan and Singapore. In our view, these exceptions are over broad and lack clear safeguards to adequately protect the copyrights of authors and publishers. Japan’s exception allows for commercial TDM activities, so long as the “exploitation” is “not for enjoying the work but merely for data analysis.” What this “non-enjoyment” requirement actually means remains unclear, despite recent efforts by the government to clarify the scope of the exception. Furthermore, rights holder opt-out is not permitted, and lawful access is not explicitly required.

The “computational data analysis” (CDA) exception under Singapore law presents similar concerns. While the Singapore exception explicitly requires lawful access, if the work copied for the TDM exercise happens to be an infringing copy or even if the infringing copy was sourced from a “flagrantly infringing online location,” this is excused so long as the user did not know or did not have reason to know that the work accessed is an infringing copy. Further, the exception makes no distinction between commercial and non-commercial TDM activities, does not permit rights holder opt-out even for commercial TDM activity, and the reproduced works (used as the training datasets) can be shared with others (for verification of the results or collaborative research).

It is worth noting that Singapore’s CDA exception may already be damaging incentives for publishers to continue investing in the business of publishing and developing the collection of works necessary to any text-and-data mining exercise, and in turn to training generative AI models. A similarly broad exception in Hong Kong would likely have the same effect — removing incentives for publishers to invest in the curation, publication, and dissemination of the high-quality works indispensable to the ethical and responsible training of generative AI models.

Chapter 5: Other Issues Relating to Generative AI — Transparency

Transparency and disclosure requirements with respect to the copyrighted content ingested by generative AI systems for training and development purposes are essential. It is in the public interest to know which works of authorship have been ingested by a generative AI model, and an essential part of ensuring proper consent is to have such information clearly recorded. Such a requirement is not burdensome and lends itself to further innovation in the field of digital rights enterprises. Virtually all stakeholders in the AI space are already subject to reporting requirements for privacy laws, for instance, and do not dispute their ability to comply with the extensive existing reporting requirements for privacy laws.

As the government notes,¹⁰ transparency is essential to the responsible and ethical development of trustworthy AI systems. Transparency promotes accountability and is necessary to build public trust in AI technologies and in the products and services into which AI technologies may be embedded.

⁹Council Directive 2019/790/EU, art. 18, 2019, [Directive \(EU\) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC](#) .

¹⁰ Consultation Paper at p. 50.

Specifically with respect to the (un-permissioned) use of copyrighted works to train generative AI systems, accurate record-keeping and appropriate disclosure of copyrighted works incorporated into training datasets will provide rights holders with the information necessary to identify the party from whom compensation is owing for such use (i.e., the training dataset creator and/or the generative AI developer). Where licensing is occurring, and as licensing models evolve, accurate record keeping of how and which copyrighted works are used to train generative AI systems may provide the mechanism through which revenue sharing models may be defined (for instance, based on how much a specific copyrighted work used as an input contributed to a particular output).

To this end, the requirements as outlined in the EU Artificial Intelligence¹¹ (AIA) are a reasonable foundation upon which to build robust transparency obligations with respect to the use of copyrighted works in the creation of training datasets for generative AI systems. Under AIA, developers of generative AI systems (or so-called frontier models) are required to “draw up and make publicly available a sufficiently detailed summary about the content used for training of the general-purpose AI model, according to a template provided by the AI Office,”¹² if the generative AI system is to be imported into/deployed in the EU market, no matter the origin of the AI model. As many generative AI systems developers have not been entirely forthcoming as to the source of the copyrighted works used to create their training datasets, such a transparency requirement is essential to provide rights holders with the information necessary to protect and enforce their rights.

A transparency requirement is also essential to determining the provenance of the content of the training datasets used to develop and train a generative AI model — traceable provenance being a building block of trust. AI systems developed or trained on works derived or created from authorized sources are more likely to yield reliable outputs than works obtained from pirated or illegal sources. It is essential to trustworthy and reliable AI that developers utilize high quality, curated content to create training corpora for their models. For example, in the case of AI training based on professional and scholarly communication, it is important that AI developers use only the Version of Record (VoR) of an article or report (appropriately licensed, of course). The VoR is the final, publisher-maintained article, updated, and archived continually in consultation with the author. Accepted manuscripts, pre-prints, or illegally uploaded text versions of the article may be subject to post publication modification or retraction, and the use of the uncorrected version could create serious and cascading scientific or medical errors in AI generated outputs.

Beyond industry economic impacts, given that AI technologies will be (and are being) integrated into applications that will impact the lives and well-being of the public, whether financially, physically, mentally, or professionally, transparency requirements are necessary to verify that high quality, peer reviewed, vetted material are used to create their training datasets, and thereby lend assurance that the output of the generative AI system trained on such vetted content is both reliable and trustworthy.

¹¹ Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act); <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689> (last accessed September 6, 2024).

¹² EU Artificial Intelligence Act, Art. 53.1 (d).

Conclusion

AAP appreciates the CEDB's consideration of the views expressed in this submission. As Hong Kong examines its copyright framework, we encourage the CEDB to ensure that human authorship remains central to and paramount in its considerations. Fostering the development of advanced technologies while continuing to protect the intellectual property rights of human authors as well as other rights holders are not inconsistent objectives, with one necessarily precluding the other. The responsible and ethical development of generative AI systems can be pursued in conjunction with respecting the rights of creators and producers, authors and publishers—without whom the creation and dissemination of the works essential to training generative AI systems would not be possible. Without authors and their literary inventions, and the publishers that bring these works to market — generative AI systems will have nothing on which to train. It should not be the government's role to promote the interests of its technology sector at the expense of its creative industries.

If we may provide further information or clarify any questions regarding this submission, please contact the undersigned. We look forward to a further opportunity to assist the government as it considers these important issues.

Sincerely,

M. Luisa Simpson

M. Luisa Simpson
Executive Vice President, Global Policy