CISAC Submission for Hong Kong's Public Consultation on Copyright and Artificial Intelligence

The International Confederation of Societies of Authors and Composers (CISAC) welcomes the opportunity to engage with the Intellectual Property Department of the Commerce and Economic Development Bureau of Hong Kong in its request for comment in the framework of its Public Consultation on Copyright and Artificial Intelligence.

CISAC is the leading worldwide organisation of authors' societies. We represent more than 5 million creators from all geographic areas and all artistic repertoires (including music, audiovisual, drama, literature, and visual arts) through our 227 members. The position of CISAC is not just a reflection of its members, but of its long history centred on defending the livelihood of creators and supporting creativity for future generations.

This submission is structured around the four main topics identified in the Public Consultation Paper (pg. 7) and aims to consolidate several key legal and policy considerations, including the possible introduction of a Text and Data Mining (TDM) exception in light of existing copyright laws.

The overall goal of this submission is to share with the Government of Hong Kong CISAC's insights into how best to preserve foundational copyright principles while, at the same time, promoting opportunities for the advancement of Artificial Intelligence (AI)[1].

Contents

---

[1] "AI" and "AI technologies" are used here as encompassing terms, including reference to foundational models (trained through various machine learning techniques), which are then adapted for downstream "AI applications" (i.e., general purpose AI systems or AI systems designed for specific purposes).

I.   Introduction

At present, a large volume of creative works is required to develop and improve AI applications designed to generate content ("generative AI"). Data containing information related to creative works, including copyright protected works, is typically gathered using so-called "data scraping" and "web scraping" tools and methods. These methods often include the use of automated webcrawling processes to gather information, sometimes applied indiscriminately and without seeking permission from rightsholders to use the content gathered for training purposes.

Thus far, the process of obtaining and using copyrighted works to train AI systems has largely gone unchecked: this is due to the lack of transparency from AI companies concerning the datasets and processes used to train their generative AI models, the lack of explicit governmental and judicial guidance on the application of existing copyright law to generative AI training, and AI companies' overreliance on the blanket assumption that such a use is permitted under copyright law or otherwise subject to an exception (i.e., a TDM exception). **As such, the current process actively disrupts the essential balance of copyright law by creating a strong bias in favour of AI technologies. This disruption threatens the livelihoods of members of the creative community, as generative AI may be used to create content that could lower, or even replace, demand for works that are the result of human creativity.**

From our understanding of the Consultation Paper, the Government of Hong Kong has a strong interest in promoting innovation, which often correlates positively with economic growth. The use of AI technologies within numerous sectors is already demonstrating enormous economic potential. At the same time, it is also important to note that the creative industry has been one of the economic pillars of Hong Kong. In 2021, the added value of arts, culture and creative industries was estimated at $124.8 billion, representing around 4.5% of the Hong Kong's GDP.[2] While we acknowledge the importance of supporting the development of AI companies in Hong Kong, a review of the copyright protection regime in Hong Kong should take into account the importance — both economic and cultural — of creators, predominantly those of Hong Kong.

Among its considerations for the updates to the existing copyright laws in light of recent technological advancements, the Government of Hong Kong is now considering the introduction of a TDM exception, which would enable developers to more freely use copyrighted works and information to train their foundational models, both in commercial and non-commercial contexts. **However, we attest that the introduction of a TDM exception would be potentially devastating to the creative and cultural sector, and may cause lasting damage to the livelihoods of domestic and foreign Hong Kong creators.** Therefore, any consideration of a TDM or TDM-styled exception (i.e., for computer analysis) should be carefully considered in light of other international positions on TDM.

Additionally, while we acknowledge that the existing laws of Hong Kong can potentially be used as a basis for determining the copyrightability of generative AI outputs created using – in whole or in part – generative AI technologies, we also find that the concepts of originality and human authorship in the creation of a work should not be dismissed, but rather considered in the interest of aligning with international developments.

---

[2] CreateHK, "Hong Kong: The Facts, Creative Industries." June 2023. Available at:
https://www.ccidahk.gov.hk/en/creative_industries_en%202.0.pdf.

II. Copyright protection of works generated by generative AI ("AI-generated works")

- **Do you agree that the existing Copyright Ordinance ("CO") offers adequate protection to AI-generated works, thereby encouraging creativity and its investment, as well as the usage, development, and investment in AI technology? If you consider it necessary to introduce any statutory enhancement or clarification, please provide details with justifications.**

As recognised in the Consultation Paper, governments around the world have adopted diverging approaches on the copyrightability of computer-generated works, creating new issues by, in some cases, classifying AI-generated works[3] as protectable subject matter under copyright law. This has resulted in a distinct divide between those jurisdictions that use human authorship and originality as key markers of protectability under copyright law, and those that define a wholly different class of protections which does not hinge on the presence of human authorship.

In 1997, the Government of Hong Kong established a comprehensive system for recognising the protectability of computer-generated works (CGW). The Consultation Paper outlines key facets of this system, including a more limited number of years of protection in comparison to works created as a result of human authorship (i.e., author's life plus 50 years versus 50 years from the creation of the work) and the entitled parties including the "necessary arranger" (i.e., the person who undertakes the arrangements necessary for the creation of the work). In further defining the "necessary arranger", the Consultation Paper analogises this concept to the idea of a "producer" within existing copyright laws in connection with sound recordings and audiovisual productions.[4]

While we acknowledge that the existing copyright laws of Hong Kong define the nature and level of protections offered to "necessary arrangers" in the course of creating a computer-generated work (CGW)[5], as we explain below, attempting to extend the interpretation of CGWs to AI generated outputs may be problematic. Not only would this position be legally novel on an international level, but the unclear distinction between purely AI-generated works and AI-assisted works may result in an unintended bias against creators in the recognition of copyright in works.

In the UK, a jurisdiction which includes CGW provisions which have served as a basis for those included in the CO, the original purpose of the legislation was to enable computer programmers and developers to protect the software they created, as well as its subsequent outputs, because they define the rules by which the software operates. Because there exists a direct and traceable causal link between the developers' definition of rules and the output a software can produce, it can be said that a CGW may be

---

[3] In this submission, we retain the same differentiation established in the Consultation Paper (pg. 8, fn. 7) between purely **AI-generated works** (i.e., works without a human author, created solely on users' prompts) and **AI-assisted works** (i.e., works created by human authors who utilise AI systems as a tool to aid their creative processes).

[4] We note that analogising the definition of "producer" in creative works to the author or first copyright owner in cases of non-human created non-LDMA works is a highly problematic interpretation of law, as it creates the possibility of AI developers (as "producers") to claim the benefits and rights in such works as "authors." (See, Consultation Paper, pg. 12: "However, in the specific case where the making of an AI-generated film does not involve a human principal director, the fair reading of the relevant statutory provisions suggests that the producer would be recognised as the author and the first copyright owner." Such a "fair reading" of this provision could also be against the legislative intention of joint authorship of a film, which should include both producers and the principal director. In such cases a "human" director is always one of the authors of a film.)

[5] Specifically, according to the Consultation Paper, AI generated works may benefit under the same legal regime as a CGW, entitling "necessary arrangers" with 50 years of protections from the creation of the work, and a moral right against false attribution of a work (pgs. 9-10). However, in the case of AI-assisted works, "…the established principles of the current copyright law are generally applicable to AI-assisted works", meaning that authors are able to benefit from protections based on the duration of an author's life plus 50 years after death, and a broader range of moral rights (pg. 8 fn. 7; pg. 9).

considered the result of the developers' creativity and thus eligible for copyright protection.[6] This is in stark contrast to how generative AI software is designed and operates, as the outputs are random, unpredictable, and varied: identical inputs may generate numerous different results. In considering this distinction, the developers of AI applications may not be eligible to benefit from protections over generative AI outputs as "necessary arrangers."

Moreover, neither the UK government nor the courts have taken a position which suggests that AI generated outputs are protectable as CGWs under existing law. As the Consultation Paper itself identifies, "the UK conducted ... a round of consultation in 2021/22 which focused on the copyright protection for computer-generated works, licensing and copyright exceptions for text and data mining .... given that a proper evaluation of the use of AI was by then impossible, and that any changes might result in unintended consequences, the UK government ultimately decided to maintain the status quo with its CGWs provisions, but would instead keep the law under review" (pg. 14). We suggest that the Hong Kong government, likewise, should be cautious about extending the interpretation and definition of CGWs to the outputs of generative AI.

The prevailing view in other jurisdictions, such as the US, Mainland China, and Japan, is that copyright legislation only offers protection to human created works. The Hong Kong Government may want to monitor the policy developments of these jurisdictions before making any attempt to extend copyright or copyright-like protections over the outputs of generative AI. This is especially the case since a diverging approach on fundamental principles of copyright law could have a significant, disruptive effect on commerce, and could ultimately weaken the international recognition works qualified as such under Hong Kong law (but not elsewhere).

Second, distinguishing the amount and scope of human involvement in the creative process in order to differentiate between different classifications of works (AI-generated vs. AI-assisted) can create a bias in the evaluation of works, as there are potentially greater benefits to be derived from works which are classified as "AI-assisted".[7] For example, in order to benefit from the classification of the work as "AI-assisted" as opposed to "AI-generated", a human author (i.e., a user of generative AI platforms) can make significant use of a generative AI model via prompt-based activity to produce an output, while adding some minimal additional creative elements expressly to change its classification. This self-reporting classification system may set up incentives for abuse by AI users, which may create further downstream issues where such works may be presented in the marketplace as works of human authorship.

- **Have you relied on the 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?**

CISAC has not directly relied on CGW provisions. However, as a general matter, CISAC, through its member societies, has been carefully monitoring the international situation related to the protectability of the outputs of generative AI. While no broad consensus has been reached yet, the interpretation of the originality requirement and, likewise, the level and extent of human involvement in the creation of a work through the use of generative AI, remains crucial in the determination of protectability of works

---

[6] See, Nova Productions Ltd. v. Mazooma Games Ltd. [2006] EWCH 24 (aff'd by Nova Productions Ltd. v. Mazooma Games Ltd. [2007] EWCA Civ. 219) (holding that the developer of the video game "devised the appearance of the various elements of the game and the rules and logic by which each frame is generated and [] the relevant computer program. [The developer] is the person by whom the arrangements necessary for the creation of the works were undertaken and therefore is deemed to be the author [of the composite images generated in-game].")

[7] See n. 4, above.

under copyright law. We maintain that – even in jurisdictions such as Hong Kong which recognises copyright-like protections in CGWs – the concepts of originality and human creative involvement may need to evolve in order to ensure that copyright laws adequately serve creators and rewards human creativity above all. At the same time, it is acknowledged that the exercise of dividing human- vs. AI-generated aspects of individual works may prove too difficult to support in the long-term. Moreover, there are odd results to consider when human creators use AI outputs and further "arrange" these materials via prompts. Under such a reading of CGW provisions as proposed, the adapted work may benefit from copyright protections while the underlying "source" AI output did not meet this requirement.

Therefore, we recommend that Hong Kong continues to monitor the international situation to continue to align with jurisdictions which will likely adapt and redefine the nature and concept of originality and/or level of human involvement necessary in recognising copyright or copyright-like protections in AI outputs.

- **Do you agree that the contractual arrangements in the market provide a practical solution for addressing copyright issues concerning AI-generated works? Please elaborate on your views with supporting facts and justifications.**

In CISAC's view, the current contractual arrangements in the market are capable of adequately safeguarding the copyright-related issues in AI-generated works. For example, some AI services have chosen to automatically assign any rights and related interests in the outputs generated by users to the service via passive license included in their Terms of Service. In such cases, taking Hong Kong's current copyright law into consideration, a user deemed a "necessary arranger" of an AI-generated work as a CGW would be stripped of all protections, commercial benefits, and moral rights (i.e., right of attribution) under such contractual arrangements.

To provide an example, as of 24 July 2024 the Terms of Service imposed by Suno[8], a prompt-based AI music generation service, includes the following provisions:

- *"Commercial Use: Subject to the Content Section below, unless otherwise expressly authorized herein or in the Service, you agree not to display, distribute, license, perform, publish, reproduce, duplicate, copy, create derivative works from, modify, sell, resell, grant access to, transfer, or otherwise use or exploit any portion of the Service, and any Output, for any commercial purposes.*

- *By using the Service or otherwise transmitting Submissions to us, you grant to Suno and our affiliates, successors, assigns, and designees a worldwide, non-exclusive, fully paid-up, sublicensable (directly and indirectly through multiple tiers), assignable, royalty-free, perpetual, irrevocable right and license to use, reproduce, store, modify, distribute, create derivative works based on, perform, display, communicate, transmit and otherwise make available any and all Content (in whole or in part) in any media now known or hereafter developed, in connection with the provision, use, monetization, promotion, marketing, and improvement of our products and services, including the Service and the artificial intelligence and machine learning models related to the Service.*

- *[a]dditional uses by Suno and other users of the Service is made without compensation to you or any other provider of the Submissions with respect to the Content, as the use*

---

[8] Suno, "Terms of Service" last accessed 13 August 2024. https://suno.com/terms/.

of the Service by you is hereby agreed as being sufficient compensation for the Content and grant of rights herein.

- *Furthermore, and for the avoidance of doubt, you irrevocably waive any and all so-called "moral rights" or "droit moral" that may exist in or in connection with the Content."*

To summarise, the owners of generative AI systems like Suno are able to leverage a distinct commercial advantage by automatically assigning and obtaining, as a condition of the use of their technology, copyright protections in the works created by users, while at the same time eliminating the ability of their users to commercially benefit from such works.[9] In our view, this style of contract which automatically assigns and/or strips any available copyright, related right, and/or moral rights interests from the creator runs afoul of long-established copyright principles. We believe that such unfair contractual practices imposed by AI service providers should be carefully monitored, and, in cases of abuse, should be sanctioned. Of course, these terms should also be monitored in the case of the creation of AI-assisted works, which shall ideally benefit from full copyright protections under Hong Kong law.

In addition, without any clear initial authorisation from authors whose works are used to train AI models, the outputs of such models may infringe on copyrighted works where such outputs replicate original elements of the source material (see Section III, below). This replication of original creative elements can occur without the knowledge of the end-user. Therefore, AI service owners and developers should be compelled to adopt transparency measures, e.g., publicly disclosing the works they have used to train their models. Finally, such AI service owners and developers must seek to obtain licenses for the use of original works for training purposes, in order to ensure that copyright is not infringed and authors are properly remunerated for the use of their works (bullet point 1, Section III).

III. Copyright infringement liability for AI-generated works

- **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? If you consider it necessary to introduce any statutory enhancement or clarification, please provide details with justifications.**

As demonstrated through the "Illustrations" provided in the Consultation Paper, the existing copyright laws of Hong Kong allow for an analysis of the possibility of infringements on a case-by-case basis, where either the operators of generative AI systems or the users may be held liable for copyright infringement depending on the specific circumstances surrounding each allegedly infringing act (pgs. 23-24). CISAC supports this balanced approach to the evaluation of potential infringements in this space, but also raises concerns regarding the assignment of infringement liability.

Specifically, CISAC identifies that paragraph 3.18 (pg. 28) does not adequately capture the reality of current market practices in the AI industry by interpreting the distribution of liability as balanced between AI system owners and end-users.[10] To provide an example, in the Terms of Service provided by

---

[9] Suno assigns its paid-tier subscribers all "right, title, and interest to and interest in and to any Output owned by Suno and generated from Submissions made by you through the Service during the term of your paid-tier subscription". However, this provision is conditional upon users' "compliance with these Terms of Service," which also includes a section on Commercial Use explicitly prohibiting the use of any part of the Service or Outputs for any commercial purposes. Suno, "Terms of Service" last accessed 13 August 2024. https://suno.com/terms/.

[10] Commerce and Economic Development Bureau of Hong Kong, Intellectual Property Department (2024). "Copyright and Artificial Intelligence Consultation Paper," pg. 28. https://www.ipd.gov.hk/en/copyright/current-topics/public-consultation-on-copyright-and-artificial/index.html. ("Additionally, the prevailing market practice of the AI industry in employing contractual arrangements for inclusion of terms of use or service between AI system owners and end-users appears to be a practical and feasible approach to address infringement liability issues associated with AI-generated works. These contractual terms facilitate a mutual understanding

Suno, users must "...represent and warrant that [they] own all right, title and interest in and to Submissions, including all copyrights and rights of publicity contained therein, and that [they] possess all necessary rights or have obtained all consents necessary to grant Suno the rights and licenses herein."[11] However, users may not be able to properly evaluate the potential use of copyrighted materials in the output of generative AI where, unknown to them, such works were used without proper authorisation in order to train the AI model. As such, due to the obscure nature of the data used to train generative AI models, users may not be adequately positioned to evaluate the potentially infringing nature of their actions when using AI models trained on infringing content.

CISAC therefore recommends that the Government of Hong Kong encourages, above all, the use of fully licensed copyrighted content to train AI models, and maintains that any responsibility and/or liability for ensuring compliance with existing copyright laws – specifically in reference to the use of protected content for AI training purposes – should be borne by AI service owners and developers.

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

One major difficulty experienced by creators who suspect that their works have been used to train generative AI models is evidentiary. The creators may find outputs produced by generative AI to feature stylistic characteristics, patterns, colour palettes, and other qualities which closely resemble or completely duplicate original, creative elements of their existing works, but may be unable to provide sufficient causal links between the availability and access to their works online, the scraping of such works from the internet, the subsequent use of such works' information to train AI models, and the appearance of unique characteristics of their original works in the outputs of generative AI. To address this challenge, CISAC strongly advocates for transparency in the data training processes of AI service owners and developers, and recommends that the operators of such services provide clear and publicly accessible information concerning the works that have been used for AI training purposes. This would ensure that existing copyright laws are fully respected in the process of advancing AI technologies.

- **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?**

CISAC disagrees that current contractual terms between AI system owners and end-users offers a concrete and practical basis for resolving disputes over copyright infringements in relation to these outputs of generative AI.

As mentioned in the first bullet point of this Section III, contractual terms which absolve AI system owners and developers from any liability of copyright infringement, and instead fully reallocate such responsibilities to its end-users, create an unbalanced and unfair system where such end-users may be unaware of the unauthorised use of copyright protected materials in the training process of the AI system.

First, as a matter of legal principle, liability is allocated or assumed by the party with the greatest degree of knowledge, or access to knowledge, of the potentially infringing act. Likewise, the owner of an AI system should assume such liability of – or at least, the responsibility for – providing its end-users with a service which has utilised legal and/or fully licensed copyrighted content in its training. This allocation

---

between AI system owners and end-users regarding their respective obligations and potential liabilities. This approach helps promote responsible and legitimate use of AI-generated works.")

[11] Suno, "Terms of Service - Content" last accessed 13 August 2024. https://suno.com/terms/.

of liability and responsibility to AI system owners is the only clear way to avoid issues of copyright infringement which may be raised concerning the outputs of generative AI.

Additionally, end-user agreements such as terms of service and terms of use are different in-kind from contracts under which terms are negotiated. Such contracts of adhesion offer "take-it-or-leave-it" terms, often highly skewed in favour of the platform or service imposing them. Therefore, it cannot be easily assumed that the contracts offered by AI platforms will provide a balanced solution for users, nor original creators.

Finally, if contractual terms are to be relied upon, stripping users of any moral rights they may be entitled to under applicable law should also be expressly prohibited (see bullet point 3 of Section II, above).

IV. Possible introduction of specific copyright exception

- **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?**

   A. The introduction of a TDM exception poses significant risks of abuse by AI developers.

During the course of many consultations on AI within the last few years, CISAC has observed that AI developers lean heavily on blanket assumptions that they are able to benefit from an exception for freely taking and using copyright protected materials, yet fail to provide specific reasoning or specific factual bases for applying such exceptions. In other words, AI developers fail to acknowledge that "exceptions," by their very nature, are exceptionally applied.

It is a longstanding principle under international law that any exception to the exclusive rights of authors over reproductions of their works must be confined to "certain special cases, provided that such [use] does not conflict with a normal exploitation of the work and does not unreasonably prejudice the legitimate interests of the author."[12] This limitation on the scope of exceptions, known as the Berne Convention "Three-step test"[13], by its very nature requires that an exception shall not be applied in an overbroad manner, to cover general categories of uses. Necessarily, copyright exceptions are recognised and applied in circumstances that do not cause undue prejudice towards a creator's ability to benefit from the uses of his or her work.

One particularly significant overgeneralization advanced by AI developers is that the presence of TDM exceptions in the national laws of certain jurisdictions *de facto* permit the use of copyrighted works for training AI. This is an incorrect interpretation of law regarding TDM exceptions and their application. In the EU, a groundbreaking Study was recently presented before the European Parliament providing proof that the reproduction of works by AI models constitutes a copyright-relevant reproduction, and that making them available on the European Union market may infringe the right of making available to the public.[14]

 As the Government of Hong Kong has already recognised (Consultation Paper, pgs. 39-42), the language, interpretation and application of TDM exceptions vary greatly between jurisdictions. Additionally, it is unclear to what extent these provisions are compatible with international treaties,

---

[12] Berne Convention for the Protection of Literary and Artistic Works (1967), Art. 9(2) (as applied to Hong Kong from July 1, 1997).
[13] See also, reference to the « three-step test » in TRIPS Agreement, Art. 13 (of which Hong Kong is party via WTO membership).
[14] Report presented before European Parliament on 5 September 2024, Initiative Urherberrecht, "Copyright & Training of Generative AI - Technological and Legal Foundations." Press Release : https://urheber.info/media/pages/diskurs/ai-training-is-copyright-infringement/2f3af96421-1725521189/240905_iu_press-release_study_ai-training-copyright-infringement.pdf.

including the Berne Convention's three-step test as mentioned above. It is therefore not possible to conclude that the existing TDM exceptions provide AI developers with a sweeping permission to use copyrighted content for AI training purposes. Even considering the potential application of another existing exception or limitation to copyright, we contend that it is in the best interest of stakeholders that the use of copyrighted materials as training data for foundational models should be licensed in all cases.

> B. *AI developers should be made responsible for clearing the information used to train their models.*

AI developers may further attempt to justify the need for a TDM exception by arguing that the task of identifying the rightsholders of works for licensing and/or training purposes is technically or commercially infeasible. This is, however, not the case for several reasons.

First, for many years, several companies have employed content recognition technologies and identification services which can automatically detect the presence of copyright-protected materials through embedded metadata and digital fingerprinting.[15] In the creation of such services, rightsholders and their affiliated organisations (i.e., collective management organisations (CMOs) and publishers) have played a significant role in disclosing vast amounts of rightsholder information to online platforms, providing timely corrections to incorrect information, and coordinating and resolving disputes where legitimate interests of follow-on creators have been at stake. Such technologies have become an industry standard in the content creation and content sharing spaces, and have become a baseline for creating reliable services which can trace authorship back to original creators. Hence, the idea that there are insurmountable technical and commercial difficulties inherent in the task of locating rightsholders for the purposes of licensing and resolving disputes is not one which can be used to justify the need for a broad TDM-styled exception.

Second, AI developers and database creators have many choices in how and where they obtain data in order to train their models and create new services. CMOs have always been proactive and open to negotiations with users. Such organisations are used to developing diverse licensing solutions to meet the very specific needs of new markets and new market players, and have historically adapted to many different business models, including in the age of large-scale industrial copying. This was the case when streaming services emerged more than ten years ago and now, almost all streaming services – including UGC platforms – are licensed and can freely use a large worldwide repertoire to attract consumers to their services. Collective management infrastructures already exist which can collect and distribute royalties, even in the face of new and complex licensing terms and limitations.

A developing AI company can make the decision to negotiate with copyright rightsholders to obtain a richer dataset to work with – and creators should be able to benefit from such innovation. Hence, with responsible data sourcing practices, innovation can thrive without the need of introducing a TDM exception.

- **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?**

The proposed TDM exception would create, rather than remove, obstacles in the enforcement of copyright and in the preservation of the creative industry. In jurisdictions without such an exception (e.g. Australia, Canada), the only option is fair licensing. This means that AI innovators, like any other digital content or service provider, must obtain licenses for the content it uses and hosts on their service. Likewise, through such licenses, creators can be assured that the uses of their works are fairly remunerated. Encouraging licences and an adequate remuneration for rightsholders will bring legal

---

[15] Youtube Help. "What is Content ID?" https://support.google.com/youtube/answer/2797370?sjid=8262792394298614957-EU.

certainty to the development of AI systems while boosting the creative industry and promoting innovation.

Furthermore, the introduction of the TDM exception that enables AI developers to avoid licensing the use of copyrighted works eliminates a necessary means for creators to benefit from the extensive use of their works in the development of AI. The effect on creators is particularly dramatic as generative AI platforms, trained on the creators' works, can then be used to produce content that would compete with the creators' works in the market for content. Where licensing opportunities are reduced, as in the case of an AI company relying on the application of a TDM exception, creators lose a crucial means of being fairly compensated for the use of the products of their creativity and labour in AI development.

- **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?**

The use of copyright-protected works by AI developers for training purposes should be considered a use subject to copyright authorisation, because it involves at least the rights of reproduction, preparation of derivative works, and distribution; and which may also implicate the right of communication to the public and adaptation rights, among others. In recognising this, fair licensing practices should be encouraged between creators and AI innovators interested in using copyrighted works as training data. Nevertheless, licensing options need to be accessible in order to be effective in this space, and organisations – such as CMOs – already exist which are well-placed to develop such solutions.

As proof that a market exists for licensing creative content for training purposes in the first place, in its response to the U.S. Copyright Office's Inquiry on Artificial Intelligence, OpenAI mentions that its models are pre-trained using "nonpublic information that we obtain from third parties through commercial arrangements", clearly indicating that a market for licensing content for training purposes has already emerged, and that AI developers are willing to negotiate and pay for access to such non-publicly available content.[16] OpenAI ultimately concluded a license with the Associated Press in July 2023 for the use of its news archive in the development of generative AI models.

CISAC is particularly well positioned to comment on the feasibility of developing new licensing and rights management solutions due to its long history supporting collective management organisations and authors' societies in developing universally accepted standards[17] for accurately tracing and tracking royalties owed to creators. While AI developers might claim that licensing works for AI training purposes is "impossible", as supported by our member societies, the collective licensing of creative works in the context of their use to train AI is a feasible option, and would provide new opportunities for collaboration and cooperation between all involved stakeholders. In fact, our members are actively in the process of developing new licensing models, some of which have been available to AI developers since 2023.

In a statement made on 12 October 2023, CISAC's member society SACEM (representing the authors, composers and publishers of music in France) announced that it had opted out[18] of the data mining of the works included in its repertoire, particularly by entities developing artificial intelligence tools.[19] As

---

[16] U.S. Copyright Office Consultation on Generative AI and Copyright, "Reply Comment of OpenAI", October 30, 2023, p. 5. Available at: https://www.regulations.gov/comment/COLC-2023-0006-8906.

[17] For example, International Standard Musical Work Code (ISWC) and International Standard Recording Code (ISRC), Interested Party Information (IPI), among numerous other technical standards currently accepted industry-wide. CISAC, "Information Services", https://www.cisac.org/services/information-services.

[18] Based on Article L122-5-3 of the French Intellectual Property Code (implementing Article 4(3) of Directive (EU) 2019/790), the provision allows rightsholders to explicitly reserve the use of their works for text and data mining.

[19] SACEM, 12 October 2023. "Sacem, In Favour of Virtuous, Transparent, and Fair AI, Exercises Its Right To Opt-Out" https://societe.sacem.fr/en/news/our-society/sacem-favour-virtuous-transparent-and-fair-ai-exercises-its-right-opt-out.

a result of this opt-out, SACEM is actively engaging with AI companies to licence the use of its repertoire for machine learning purposes.

Such licensing options need to be widely adopted to ensure a sustainable future for creators and innovators alike, and can be effectively fostered through existing infrastructures used for managing rights which are already present across all creative sectors.

- **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?**

While we strongly maintain that a TDM exception should not be introduced, we include several necessary considerations below to aid the Hong Kong Government in their legislative process, should they consider the adoption of such an exception.

A. *Any TDM Exception introduced should be limited to non-commercial purposes only.*

If Hong Kong's proposed TDM exception is expanded in scope to include commercial purposes, there is a significant additional risk that copyright laws can be undermined or circumvented. The non-commercial purposes limitation serves to provide opportunities for learning for specific and narrowly-defined research purposes, non-profit projects, and other instances where the public interest outweighs the need for obtaining a license and remunerating creators. This limitation further ensures that creators are not deprived of the opportunity to be fairly paid for their creative work in cases where there is no compelling public interest at stake.

B. *Any TDM Exception should be accompanied by a compensation mechanism.*

We maintain that any TDM exception introduced into Hong Kong law should nevertheless include (in the event it is not subject to a rightsholder opt out) a means for rightsholders to be compensated for the use of their works for AI training purposes. The commercial impact of the use of original creative works for training purposes is twofold: when AI models produce works that are substantially similar and/or replicate original creative elements or stylistic features of original creators' works, such AI-generated outputs can create direct competition in the market against original works, while the amount and scale of AI output activities may act to further dilute it.

C. *Any TDM Exception should provide clear technical standards concerning an "opt-out" protocol for rightsholders.*

So far, national governments throughout the world that have introduced a TDM exception have failed to articulate clear technical standards on the exercise of rightsholders' opt-out, which has caused diverging industry practice and an unclear international legal landscape overall. While several rights management organisations have attempted to exercise the "opting-out" of processes which gather their rightsholders' works and information for AI training purposes,[20] the effect of this approach on the preservation and enforcement of creators' rights has not been clear. Additionally, simple technical measures such as the addition of "robots.txt" files on webpages to deter webscraping practices has not been completely effective, as some companies may use webcrawlers which either ignore or circumvent such technical measures in order to scrape data without any repercussions. CISAC recommends that this situation should be anticipated by the Government of Hong Kong, and that if it considers the introduction of a TDM exception, that it also incorporates clear and accessible means for rightsholders to successfully opt out of the use of their works for training purposes, while ensuring that

---

[20] See, e.g., SACEM, 12 October 2023. "Sacem, In Favour of Virtuous, Transparent, and Fair AI, Exercises Its Right to Opt-Out" https://societe.sacem.fr/en/news/our-society/sacem-favour-virtuous-transparent-and-fair-ai-exercises-its-right-opt-out.

such a technical standard and/or protocol is designed to enable AI developers to refrain from the use unauthorised works for AI training purposes.

   D. Any TDM Exception should be accompanied by transparency obligations for AI developers.

AI developers shall be required to disclose the relevant information about the copyrighted works used for training purposes (pg. 48), even if a TDM exception applies. Transparency obligations in the disclosure of data sets will enable rightsholders to better understand when their works are used in compliance with the applicable legal framework (e.g. within the limited scope of a TDM exception), and where a license should be applied.

   E. Any TDM Exception should be accompanied by sanctions against abusive data mining conduct on behalf of developers.

Rightsholders are currently struggling to safeguard their content in online spaces where the opt-out alone has not provided a sufficient safeguard against the abusive practices of AI developers. For example, AI startup Anthropic has recently been accused of "egregious" data scraping practices, where the service's web crawler hit the servers of several websites millions of times within the span of a few hours, generating server stress and website malfunctions among other damages to the website owners.[21] The website owners may even include in the terms of services a prohibition against the use of its data for machine learning, but such prohibitions may not be enough to prevent the AI service from other damaging conduct. Therefore, a TDM exception should not only be narrowly defined and applied, but should also be enforced alongside other measures which prevent developers from engaging in abusive conduct targeted towards copyright owners.

We therefore advise that, in the interest of promoting a fair system which safeguards the interests of original creators, that AI service owners who seek to benefit from a TDM exception should likewise remunerate authors for the use of their works for training purposes.

V.   Other issues relating to generative AI

As a final comment concerning the broader aspects of generative AI not specifically concerning copyright, related rights or moral rights protections, particularly those concerning the use of so-called "deepfakes"[22] and other means of using personal information for fraudulent purposes, CISAC recommends that the Government of Hong Kong clarify that copyright exceptions, such as the parody exception under Hong Kong law[23], cannot be applied to justify the creation and use of "deepfakes" or other fraudulent representations of real individuals through the use of AI.

VI.   Conclusion

CISAC firmly believes that a more robust creative marketplace for human creators can coexist with, rather than undermine, innovation in the field of AI. Richer and more diverse data sources are the key to the development of better AI models, and the promise of copyright provides human authors with the ability and incentives to make a living from their works. Once creators are able to fully benefit from uses of their works, the promise of copyright can continue to be kept in an AI-driven era.

Overall, the Government of Hong Kong should monitor the international situation before considering any broadening of the scope and interpretations of existing law. Considerations for the introduction of

---

[21] Hammond, George (2024). "AI start-up Anthropic accused of 'egregious' data scraping." Financial Times Online, 26 July 2024. https://www.ft.com/content/07611b74-3d69-4579-9089-f2fc2af61baa.

[22] Deepfake refers to the combination of the concepts "deep learning" and "fake" to describe the utilisation of AI or deep learning algorithms to create believable or realistic videos, images and audio.

[23] Copyright Ordinance (Hong Kong), Division III, Acts Permitted in Relation to Copyright Works, 39A. "Parody, satire, caricature and pastiche". https://www.elegislation.gov.hk/hk/cap528?pmc=0&m=0&pm=1&SEARCH_WITHIN_CAP_TXT=pastiche.

new exceptions into law should also be carefully weighed to ensure they do not unduly harm existing creators' rights or disrupt the balance between private rights and public interests. Introducing a new TDM exception into Hong Kong law raises serious concerns about potential breach of international treaties and fundamental principles underlying copyright law, which are in place to ensure that authors can properly benefit from the uses of their works.

Any pre-determined position of the Hong Kong government favouring one industry should be avoided. Ultimately, CISAC strongly recommends against the adoption of new TDM exceptions that permit AI systems to commercially exploit copyrighted works without guaranteeing, via license, rightsholders' explicit authorization and remuneration, among other reasonable safeguards such as transparency obligations.

CISAC once again appreciates the opportunity to provide comments on the Government of Hong Kong's Public Consultation on Copyright and Artificial Intelligence and looks forward to continuing this important international dialogue to safeguard the future of creators.