

1. Introduction

The Asia Internet Coalition (AIC) welcomes the Government of Hong Kong's deep commitment to encouraging creation and investment in creativity while supporting innovation, as highlighted by Hong Kong's objective under the National 14th Five-Year Plan to develop Hong Kong as an international I&T center and an East-meets-West center for international cultural exchange. Our members are among the leading innovators in the burgeoning area of Artificial Intelligence (AI) and continue to work to harness the power of AI for the betterment of all humanity. Given the profound implications for the future development and deployment of transformative AI innovations, AIC appreciates the Bureau's careful consideration of the legal framework that will govern this technology. Copyright is an essential component of that legal framework calibration, and we welcome the opportunity to contribute to this discussion.

We believe AI can contribute to a world where language barriers evaporate and people can communicate seamlessly in real time; where governments, the private sector, and civil society can come together to find solutions to the challenges of climate change before it's too late; where vaccines can be developed and deployed before a pandemic takes grip of humanity; and where a region's cultural and language are deeply understood, respected, and form a part of our modern day technical fabric. However, Hong Kong could miss these opportunities if the delicate balance between copyright, innovation, and competition is not aligned. As industry leaders, we appreciate the opportunity to provide our views on the issues raised in the Hong Kong Commerce and Economic Development Bureau Intellectual Property Department Copyright and Artificial Intelligence Public Consultation Paper. Specifically, these comments seek to address two key topics posed by the Hong Kong paper: introduction of a text and data mining (TDM) exception and copyright infringement liability for AI-generated works.

AI and Generative AI technology is still nascent, but is already dramatically changing the way people globally are able to access information, create content, perform tasks, conduct groundbreaking research, and enjoy entertainment.

2. Proposed TDM Exception

AIC supports the proposed TDM exception to Hong Kong's copyright law. Adopting such an exception would help establish a positive, pro-innovation precedent and that could profoundly impact the development of AI technology in Hong Kong. This exception would clearly allow developers and researchers to access and analyze data, including copyrighted material, critical for the purpose of TDM. Currently, Hong Kong's copyright law does not expressly permit TDM, which can lead to uncertainty and reluctance from industry leaders to invest in AI research and development in Hong Kong. The proposed exception would provide much-needed clarity and certainty, encouraging investment in AI technology and fostering innovation. This, in turn, could accelerate the development of wide-ranging AI applications, from natural language processing and machine translation to medical diagnosis and drug discovery. Moreover, adopting the proposed TDM exception would be consistent with similar, pro-innovation exceptions in place in Japan and Singapore. Finally, AIC also supports the Bureau's proposal that the proposed exception not be restricted to non-

commercial research and study because, the Bureau noted, much of the innovation in the AI ecosystem that should be incentivized is by commercial entities.

Comments on broad and flexible TDM exception

- a. We support the introduction of a broad and flexible TDM exception in the Copyright Ordinance CO, which could help in driving the development of AI in and transforming industries across Hong Kong. AI models, especially foundational large language models, require access to large, diverse datasets to be effectively trained and reach their full potential. Without a broad and flexible TDM exception, AI developers face significant uncertainty and potential barriers and transaction costs, hampering their ability to build cutting-edge AI systems. Moreover, overly restrictive copyright regimes that impose hurdles to the use of publicly available data for AI training will distort the competitive landscape, stifle the emergence of innovative AI startups, and increase the risk of inaccurate or biased models being released. Therefore, a broad and flexible TDM exception that allows the use of copyrighted works for commercial and non-commercial TDM activities would provide AI developers the necessary legal certainty and flexibility to utilize a wide range of data sources, which is crucial for securing Hong Kong's position as a leading international innovation and technology hub.
- b. A broad and flexible TDM exception is also consistent with the goal and rationale of copyright protection. Copyright law is fundamentally focused on protecting the unique creative expression of authors and fostering creativity for the betterment of society, not restricting access to the underlying facts, data or ideas contained within copyrighted works. Copyright is intended to incentivize the production and dissemination of original expressive creation. A pro-innovation and flexible TDM exception upholds this principle by allowing flexibility in terms of the use of copyrighted materials for TDM activities¹. By facilitating broad access to diverse data sources, including copyrighted works, the TDM exception enables AI developers to uncover valuable insights, trends and patterns – the type of non-expressive information that copyright law was never intended to monopolize. This allows for the development of innovative AI applications that serve the public interest, while still preserving economic incentives for authors to create original, expressive works. Maintaining this balance is crucial for promoting

¹ Separately, it is important to note that an AI model does not “store” or contain copies of the data itself. Rather, the model uses its training data to guide the process of gradually adjusting its representations of abstract meaning. The resulting outputs are not determined by any one piece of training data; instead, they reflect, for example, the influence of patterns informed by up to billions of examples. This has two critical implications: a) no individual piece of content has a particular influence over the model; b) while a piece of data may assist a model in better understanding the meaning and interrelationship of words, the model does not copy the particular training data, but rather, reflects the patterns it has learned based on the training set as a whole. Thus, a TDM exception better reflects how AI models and their training works.

technological progress and ensuring a vibrant, balanced copyright system that fulfills its purpose.

- c. To maintain its position as a leading innovation and technology hub within Asia, Hong Kong should consider adopting a broad and flexible TDM exception similar to the approaches taken in neighboring jurisdictions like Singapore and Japan. As summarized in the Consultation Paper, both Singapore and Japan have adopted TDM exceptions which permit both commercial and non-commercial computational data analysis and processing of copyrighted work, with very limited conditions. In particular, Singapore's TDM exception cannot be overridden by contract terms, ensuring AI developers and researchers have legal certainty to access and utilize diverse data sources needed to train advanced models efficiently. If Hong Kong adopts a restrictive approach instead that results in AI model creators facing hurdles to developing models because of limited access to training data including copyrighted materials, those AI model creators may move AI training workloads to overseas. This will create not only a missed opportunity for Hong Kong to promote AI development and deployment within the jurisdiction, but also give rise to a risk that models have not adequately taken Hong Kong context and culture into account (due to insufficient open training materials available in Hong Kong). Therefore, aligning Hong Kong's approach with regional best practices would cement its attractiveness as a hub for AI innovation, sustain its competitiveness in the fast-evolving global technology landscape, and help ensure that models fairly represent Hong Kong's unique culture and heritage.
- d. Section 38 of the CO already provides a fair dealing exception for research activities, which AI developers could potentially rely on to conduct TDM activities using copyrighted works. TDM is the process of computational analysis to identify patterns within large data sources. The use of fair dealing exceptions for commercial TDM is also common practice in other jurisdictions. As the existing fair dealing exception in Hong Kong has a broad scope with limited conditions, we recommend that any new TDM-specific exception follow a similar approach and should not be more restrictive than the existing fair dealing exception. This would ensure regulatory consistency, where the same risk and activity is dealt with in a consistent manner. It would also avoid creating legal uncertainty that could disrupt ongoing commercial TDM activities that may have been relying on the fair dealing exception. Aligning the TDM exception with the existing fair dealing framework would provide AI developers the necessary legal clarity and flexibility to fully leverage copyrighted data sources for research and innovation.
- e. AIC recommends that the proposed TDM exception not be accompanied by conditions that significantly limit its scope and effectiveness. With respect to opt outs, AIC notes that most developers of AI models are voluntarily enabling content owners to opt out from AI training via certain industry-standard protocols. AIC believes that such existing market practices can obviate the need for a

specific opt out condition on the proposed TDM exception. However, due to the huge scale and technical impracticability of examining individual items of content, or even groups of content, standard machine-readable (i.e., coded) indicators for opting out are a critical element of any workable opt out scheme. Therefore, should an opt out system be deemed critical, at minimum, it should encompass these existing or industry-led technical, machine-readable methods of enabling content owners to opt out and avoid creating new, inconsistent requirements.

Comments on Licensing

With respect to licensing, a *compulsory* licensing solution would be counterproductive and impractical, and, unlike the proposed TDM exception, would harm innovation in the AI ecosystem. Given the massive quantity of content and the large diversity of content needed to train AI models, it would be virtually impossible for AI developers to locate the owners of all such content and administratively impossible to negotiate licenses with each of them.^[1] However, individual companies may choose to leverage licensing agreements with specific entities in order to deepen partnerships or for discrete usage of data such as that not otherwise available. We are already seeing such *voluntary* licensing agreements emerge, even in jurisdictions with TDM or TDM-like exceptions to copyright law, demonstrating that adopting a TDM exception would not foreclose similar voluntary agreements. This type of commercial decisions (i.e. opt-out and licensing schemes) should remain voluntary after the introduction of the TDM exception.

Commercial decisions, such as opt-out and licensing schemes to remain voluntary is crucial for maintaining a balanced ecosystem between content creators and data miners.

3. Limit the conditions to the TDM exception

- a. The Hong Kong Government should limit the conditions for triggering the TDM exception, in particular, the TDM exception should not be subject to an opt-out or to licensing conditions that would undermine the utility of the TDM exception. As submitted above, a broad and flexible TDM exception not only promotes the development of AI in Hong Kong, but is also consistent with the goal and rationale of copyright protection. Restrictions such as opt outs and mandatory licenses limit the opportunity for AI developers to develop and deploy the most useful and capable models. Therefore, imposing any such restriction must therefore be weighed against the societal cost of less useful and capable models as well as the disincentives to developers to do their work in HK. Rather than promoting innovation and creativity, licensing schemes of copyrighted works for TDM activities would impose significant transaction costs and administrative burdens on AI developers, particularly smaller startups and researchers, which

lead to a distortion of the competitive landscape and stifling the emergence of innovative AI startups.²

- b. Subjecting the TDM exception to opt-out or licensing scheme conditions would also fragmentize the comprehensive datasets needed to develop high-performing AI models, as not all AI developers would have equitable access to all available data sources for training their AI models. AI developers would be forced to patch together incomplete, biased datasets. This fragmentation undermines the ability to uncover the rich insights and patterns hidden within diverse, large-scale data sources - the very purpose of TDM. This can lead to suboptimal model performance, propagation of biases, and missed opportunities for breakthrough discoveries and innovations.
- c. Footnote 70 on page 31 of the Consultation Paper says that “the computational data analysis and processing activities conducted for training AI models in the model market may be of a commercial nature, and may copy and *store* the whole of a copyright work” (emphasis added). We would like to clarify that AI models do not actually store the training data used to create them. This is elaborated in a discussion paper issued by The Hamburg Commissioner for Data Protection and Freedom of Information of Germany recently³. During the training process, highly abstracted mathematical representations called embeddings are calculated from the original text data. These embeddings capture patterns and relationships between linguistic elements, but do not retain the specific content from the original training data. The trained model only stores these statistical representations, not the raw training data itself. It reinforces our argument that the use of copyrighted works for TDM activities do not involve the enjoyment of the author’s expression in the copyrighted works, and so no condition should be imposed on the TDM exception.
- d. We therefore recommend the Hong Kong Government to adopt the same flexible approach to TDM exception as Singapore, which adopts a broad and flexible TDM exception and prohibits the use of contractual terms or opt-out conditions to override the TDM exception. This could help in establishing a favourable legal regime for the development of Hong Kong into a major AI and technology hub within the region.
- e. If any conditions need to be included within a TDM exception, they should be narrowly crafted to avoid undermining the core purpose of the TDM exception. As an example, Singapore's TDM exception only applies if users have "lawful access" to the copyrighted works. However, even this seemingly straightforward condition has faced reconsideration, with the Singapore government recently launching a public consultation on whether prohibitions on circumventing

² Martens, B. (2024) “Economic arguments in favour of reducing copyright protection for generative AI inputs and outputs,” Working Paper 09/2024, Bruegel, [Economic arguments in favour of reducing copyright protection for generative AI inputs and outputs](#)

³ Hamburg Commissioner for Data Protection and Freedom of Information. (2023). Discussion paper: Large language models and personal data. [Discussion paper: Large language models and personal data.](#)

technological access controls would impair or adversely affect the dealing with copyrighted works that would be non-infringing based on permitted uses⁴. Any additional conditions, such as an opt-out right for copyright holders, must be designed with similar caution. If an opt-out mechanism is included, the intention to opt out should be readily identifiable in a clear, standardized machine-readable manner. This would minimize the administrative burden and legal uncertainty AI developers face in ascertaining the copyright status of each work they seek to analyze. It may also be worth examining limitations on how the works are used beyond TDM, such as limiting further distribution of the works. Overall, the guiding principle should be to craft TDM exception conditions that unambiguously enable access to comprehensive datasets, rather than imposing hurdles that fragment the available training data and distort the competitive landscape.

4. Copyright infringement by AI-generated works

- a. We agree with the general proposition in the Consultation Paper that the responsibility of any copyright infringement by AI-generated works should be determined on a case-by-case basis and who should be liable for such infringement is indeed fact-dependent. Nevertheless, when determining the responsibility for copyright infringement by AI-generated works, the Hong Kong Government should take into account the fact that an AI system developer does not have direct control over how the AI system responds to individual user prompts, particularly where the user is attempting to trick the system into generating an infringing output. Indeed, many potentially infringing outputs are the result of intentional, technically-designed attacks on such systems, requiring precise knowledge of the model's structure and often the content used to train the model. Many system developers now include terms in the model's terms of use that prohibit such attacks, and other efforts to trick the model into producing undesirable outputs. Under the current CO, however, there remains a possibility that under the above scenario, the AI developer could arguably still be held liable/jointly liable for copyright infringement or for authorizing an infringing act under the theory that the AI developer has (through the AI system) arguably collaborated with the AI user, or has control over the AI system that allegedly copied the whole or substantial part of a copyrighted work in producing the output. To fully support the intent of the TDM exception and give legal certainty to AI developers under that exception when copyrighted work is legitimately used, we recommend the Hong Kong Government clarify the CO such that the only party at risk for being held liable for generating an infringing output is that party who intentionally causes the prompting of the model to generate the infringing output. In addition, the Hong Kong Government should consider

⁴ Ministry of Law and the Intellectual Property Office of Singapore. (2024). 2024 Public Consultation on Prescribed Exceptions In Part 6, Division 1 of The Copyright Regulations 2021 (https://www.mlaw.gov.sg/files/2024_Public_Consultation_on_Prescribed_Exceptions_in_Part_6_Division_1_of_the_Copyright_Regulations_2021.pdf) 2024 Public Consultation on Prescribed Exceptions In Part 6, Division 1 of The Copyright Regulations 2021

including a safe harbour provision in the CO to limit AI developers' liability in circumstances where users prompt and AI developers take reasonable steps to limit or stop copyright infringement when notified (for example, notice and take down of infringing content). The prompt could cause the system to either (1) generate a facsimile of an element of its training data, or (2) generate a facsimile of some element of data that the model has not previously been exposed to.

- b. AIC agrees with the Bureau that an overly prescriptive approach towards liability for any potentially copyright infringing AI-generated works would “hinder AI technology development and undermine efforts to encourage its use”, and that contractual arrangements are an effective means of addressing any liability issues. Many companies have user agreements via Terms of Service that stipulate where liability lies and that the end user is responsible for ensuring their use complies with the law. As the Bureau notes, “[t]hese contractual terms facilitate a mutual understanding between AI system owners and end-users regarding their respective obligations and potential liabilities.”

5. Reconsider copyrightability of purely AI-generated works

- a. The Consultation Paper distinguishes between AI-generated works created and generated by generative AI without a human author based on users' prompts and works created by human authors who utilize AI systems as a tool to aid their creative processes (“**AI-assisted works**”), and AI-assisted literary, dramatic, musical, and artistic (“**LDMA**”) works are considered as ordinary LDMA works (rather than computer-generated LDMA works) under the CO. However, the current CO lacks clear guidelines on how to reliably differentiate these two categories of works. This ambiguity risks introducing significant legal uncertainty around which copyright regime should apply to a given work. We therefore recommend the Hong Kong Government to make reference to the Japanese framework, which sets out factors for determining whether AI has been used as a genuine creative tool, versus cases of purely autonomous machine generation. Specifically, the Japanese approach evaluates whether the human user had a demonstrable “creative intention” and made a substantive “creative contribution” beyond simply providing basic prompts or instructions.
- b. We urge the Hong Kong Government to carefully reconsider whether purely AI-generated works should be eligible for copyright protection. While the Consultation Paper proposes that the existing “computer-generated works” provisions in the CO could potentially cover AI-generated **LDMA** works, there are significant uncertainties around how the current copyright regime could apply in such cases. According to the CO, all LDMA works must satisfy the originality requirement for copyright to subsist, but it is highly unclear how a work generated entirely by an AI system, without any direct human authorship, could meet this criterion. Additionally, under the CO, the “author” of a computer-generated work is defined as the person “by whom the arrangements necessary for the creation of the work are undertaken”. However, with purely AI-generated works, there may

be multiple parties involved - including deployer, and user - making it ambiguous who the true "author" is. As generative AI capabilities advance exponentially, leading to a proliferation of purely AI-generated LDMA works, it is imperative that the Hong Kong Government provides clear legal certainty for all stakeholders. And we recommend that purely AI-generated works should not be granted copyright protection for the reasons below.

- c. From a policy perspective, purely AI-generated work should not be granted copyright protection as the core rationale behind copyright protection is to incentivize human creativity and expression, not to grant monopolistic control over non-expressive content generated by machines. Extending copyright to purely AI-generated works would run counter to the fundamental purpose of the copyright system. Generative AI models do not require the promise of exclusive rights to spur their development - their outputs are produced through automated computational processes, without any meaningful creative contribution from human authors. From an economic perspective, there is also no need to protect purely AI-generated work as the marginal cost of producing such outputs is very low, close to the marginal cost of reproduction, which eliminates the rationale for copyright as an incentive mechanism.⁵ The approach taken in Japan's copyright framework provides a compelling model. In Japan, materials autonomously generated by AI (i.e., material that is generated by AI without any instructions from humans or only by giving simple instructions as prompt) are not considered as creatively produced expressions of thoughts or sentiments and are therefore not considered copyrighted works. Therefore, our recommended approach is to maintain copyright's focus on original human expression, while allowing the unimpeded use of AI to generate non-expressive content that serves the broader public good.

⁵ *Supra* note 1