February 23, 2022

Submission of BSA | The Software Alliance to the Commerce and Economic Development Bureau

Updating Hong Kong's Copyright Regime: Public Consultation

Submitted via Email to: co_consultation@cedb.gov.hk

BSA | The Software Alliance (BSA) welcomes this opportunity to provide comments to the Commerce and Economic Development Bureau (CEDB) in response to the Public Consultation Paper on Updating Hong Kong's Copyright Regime (Consultation).[1] BSA is the leading advocate for the global software industry before governments and in the international marketplace.[2] Our members are at the forefront of software-enabled innovation that powers the global economy and helps businesses in every industry compete more effectively. BSA is therefore acutely aware of the critical role that copyright policy plays in fostering research and development of cutting-edge technologies and we appreciate the opportunity to provide input as CEDB considers potential reforms to the Copyright Ordinance.

This Consultation arrives a timely moment. Advances in AI and software-enabled data analytics are fueling job and economic growth in Hong Kong, improving how businesses in every sector operate, and producing real societal gains. Across industries, the analysis of data has made businesses in Hong Kong more agile, responsive, and competitive, boosting the underlying productivity of many key pillars of the economy. More than simply benefitting

---

[1] https://www.ipd.gov.hk/eng/intellectual_property/copyright/Consultation_Paper_on_Copyright_Eng.pdf

[2] BSA's members include: Adobe, Alteryx, Atlassian, Autodesk, Bentley Systems, Box, Cisco, CNC/Mastercam, DocuSign, Dropbox, IBM, Informatica, Intel, MathWorks, Microsoft, Okta, Oracle, PTC, Salesforce, SAP, ServiceNow, Shopify Inc., Siemens Industry Software Inc., Splunk, Trend Micro, Trimble Solutions Corporation, Twilio, Unity Technologies, Inc., Workday, Zendesk, and Zoom Video Communications, Inc.

from the adoption of AI, innovators in Hong Kong are on the leading edge in shaping the development of the technology.

We agree with CEDB that the Copyright Ordinance should be updated to account for technological transformations and ensure that Hong Kong business are well positioned to "leverage innovation and creativity to drive economic growth." To meet these objectives, we encourage CEDB to support the introduction of a specific copyright exception to provide clarity for organizations engaged in the development and adoption of AI technologies. There is an emerging international norm that reproductions created as part of the machine learning process should be subject to an explicit copyright exception for "text-and-data minging" (EU), "data analysis" (Japan), or "computational data analysis" (Singapore). Regardless of the chosen terminology, CEDB should move swiftly to provide Hong Kong with a solid legal foundation to support AI innovation.

<u>AI and Copyright: The Source of Legal Uncertainty</u>

The incredible advances in AI capabilities in recent years have been enabled by a particular subset of the technology referred to as "machine learning." At its most basic, machine learning involves the computational analysis of large amounts of data (i.e., "training data") to identify correlations, patterns and other metadata that can be used to develop a "model" capable of making predictions based on future data inputs. For instance, GitHub recently used machine learning to create CoPilot, an AI-powered software tool that enables programmers to write code more efficiently by providing source code recommendations for common software tasks.[3] Much like a smartphone or email "autocomplete" recommendation, CoPilot works by analyzing the code that a programmer is working on and suggesting options for completing the identified function. The model that powers GitHub Copilot was "trained" by analyzing patterns and correlations from a large repository of publicly available open source code. Tools such as CoPilot promise to democratize the software development process, empowering more people, more businesses, and more industries to benefit from the creation of customized software solutions.

As the foregoing example demonstrates, some forms of machine learning rely on training data that is derived through the computational analysis of items potentially subject to copyright protection. This "input" stage of the machine learning process may involve two sets of reproductions that potentially implicate the Copyright Ordinance: (1) reproductions necessary to create a corpus of "training data," and (2) transient reproductions that are incidental to the computational process of training the AI model. In each case, the reproductions are "intermediate" in the sense that they are not visible or otherwise made available to the public. Instead, the reproductions are the necessary byproduct of a

---

[3] https://copilot.github.com/

technical process that is aimed at identifying non-copyrightable information *about* the underlying corpus of works – i.e., the correlations and patterns that inform the creation of the AI model and enable it to make predictions based on future data inputs.

Such intermediate, non-expressive reproductions have no impact on the economic interests that copyright is intended to protect. Recognizing that not all uses of copyrighted works should require permission, the Copyright Ordinance includes exceptions that arguably might cover certain forms of information analysis and machine learning. However, because the Copyright Ordinance currently lacks an express exception to enable computational analysis, there is uncertainty about the scope of activity that is permitted under current law. To resolve the existing ambiguity, CEDB should introduce an exception that reflects emerging international norms.

AI and Copyright – International Developments

In contrast to Hong Kong, the copyright laws of several other leading AI nations provide greater legal certainty for AI research and development. Indeed, there is an increasing global awareness about the need to modernize copyright laws to facilitate the development of AI.

1. *Japan ("Data Analysis" Exception)*

Japan first recognized such a need in 2009 when it amended its Copyright Act to create an explicit exception for reproductions that are created as part of an "information analysis" process.[4] Although the 2009 amendment is heralded as having transformed Japan into a "machine learning paradise,"[5] the Japanese Diet made further revisions to the Copyright Act in 2018 to further expand the exception.[6] In May 2018, the Diet passed the Copyright Law Amendment Act, broadening the existing exception to allow users to "exploit" any

---

[4] *See* Copyright Law of Japan, Article 47-7, "To the extent that it is considered to be necessary, it is permissible to record a work onto a recording medium or to make an adaptation of a work (including recording a derivative work created by adaptation) if the purpose of doing so is information analysis (meaning the extraction, comparison, classification, or other statistical analysis of language, sound, or image data, or other elements of which a large number of works or a large volume of data is composed; the same applies hereinafter in this Article) by means of a computer; provided, however, that this does not apply with regard to database works compiled for use by persons who carry out information analyses."

[5] *See* "Machine Learning Paradise" (Tatsuhiro Ueno), available at https://rclip.jp/2017/09/09/201708column/ (Japanese)

[6] Evolving Paradise for Machine Learning—Revisions to the Copyright Act Further Accelerate Development of Japan's AI | STORIA LAW, available at https://storialaw.jp/en/service/bigdata/bigdata-12.

copyrighted work for non-consumptive purposes, including for "data analysis (meaning the extraction, comparison, classification, or other statistical analysis of the constituent language, sound, or image data)" and "computer data processing."[7] In addition to creating a general purpose exception for non-consumptive uses of copyrighted works, the recent amendment package also authorizes beneficiaries of the information processing exception to make limited public uses of the underlying works, such as the display of snippets.[8]

2. United States (Fair Use)

In the United States, courts have confirmed that under the "fair use" doctrine, incidental copies of a work made in the course of informational analysis are non-infringing, even where the analysis is performed for commercial purposes. Creating a corpus of AI training data fits neatly within a long line of fair use rulings. For instance, in *Authors Guild v. HathiTrust* and *Authors Guild v. Google*, the Second Circuit determined that the unauthorized copying of tens of millions of books for the purposes of creating a searchable database of those works was a fair use.[9] Notwithstanding Google's commercial motivation, the Second Circuit determined that the creation of the database was a fair use because it served a "highly convincing transformative purpose" that did not create "substitute competition" for the works included in the database.[10] The court reasoned that creating the searchable database was transformative because it enabled users to uncover factual information "about" the works included in the database and that providing access to such information did not implicate the expressive interests that copyright is intended to protect.[11] This conclusion reflects a strong consensus among the courts, which have consistently ruled that unauthorized copying is permitted when it is undertaken for non-expressive purposes[12] or to identify non-copyrightable information about the copied works.[13]

---

[7] Copyright Act of Japan, Article 30-4, *available at* https://wipolex.wipo.int/en/legislation/details/20024

[8] *Id.* at Article 47-5.

[9] *See* Authors Guild v. HathiTrust, 755 F.3d 87 (2d Cir. 2014); Authors Guild v. Google, Inc., 804 F. 3d 202 (2d Cir. 2015).

[10] Authors Guild v. Google, Inc. at 219.

[11] *Id.* at 217. Because "an author's derivative rights do not include an exclusive right to supply information…about her works," the Second Circuit rejected plaintiff's argument that Google's unauthorized copying "usurped their opportunity" to license access to such non-copyrightable information.

[12] *See* A.V. *ex rel.* Vanderhye v. iParadigms, LLC, 562 F.3d 630 (4th Cir 2009) (holding that reproductions that are "not related to the creative core of the works" are a fair use.).

[13] *See*, *e.g.*, Sega Enterprises Ltd. v. Accolade, Inc., 977 F.2d 1510 (9th Cir. 1992); Sony Computer Entertainment, Inc. v. Connectix Corp., 203 F.3d 596 (9th Cir. 2000)

3. European Union ("Text-and-Data Mining" Exception)

The European Union also recently passed legislation to provide clarity for the development of AI. In April 2019, the European Council formally adopted the Directive on Copyright and Related Rights in the Digital Single Market. Articles 3 and 4 of the Directive create two broad exceptions that authorize AI researchers to make reproductions that are needed for the purposes of carrying out "any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations." Importantly, the Directive clarifies that Articles 3 and 4 are without prejudice to existing exceptions and limitations that may already allow for reproductions that are necessary for machine learning. The Article 4 exception specifically contemplates that text and data mining may be performed for commercial purposes, but it is limited to circumstances where a rights holder has not "expressly reserved" his or her rights "in an appropriate manner, such as machine-readable means in the case of content made publicly available online."

4. Singapore ("Computational Data Analysis" Exception)

In September 2021, Singapore adopted important new legislation to update the country's Copyright Act and ensure that it remains a global hub for AI research and innovation.[14] Singapore's new law includes an important exception for "computational data analysis" that will provide much needed certainty for the researchers and developers fueling Singapore's AI ambitions.[15] The scope of the exception includes reproductions that are necessary for the purpose of performing a computational data analysis[16] and communications to the public[17] that are necessary for the purposes of: (i) verifying the results of the computational data analysis or (ii) collaborative research and study relating to the purpose of the computational data analysis.[18] The Singapore exception is subject to a number of important safeguards. Most importantly, it applies only to circumstances in which an entity has "lawful access" to the copy on which computational analysis will be performed, and the exception cannot be invoked if the entity performing the computational data analysis knows either that the source material is infringing or that it was "obtained

---

[14] https://sso.agc.gov.sg/Acts-Supp/22-2021/Published/20211007?DocDate=20211007

[15] https://asiaiplaw.com/article/computational-data-analysis-exception-in-singapores-copyright-act-2021-a-game-changer

[16] Division 8, Section 233(1)(a).

[17] Division 8, Section 233(1)(b).

[18] Division 8, Section 233(c).

from an online location that is being or has been used to flagrantly commit or facilitate rights infringements."[19]

Singapore's proposal is consistent with elements of forward-looking copyright policies that are important to the development of AI. The proposal is technologically neutral, applying broadly to any reproduction that is needed in order to carry out a "computational data analysis" of a work, including copies made during (and in preparation for) training an AI system. Importantly, the exception also permits the exchange of datasets to facilitate collaborative research. Recognizing that AI research is being driving by a large ecosystem of researchers and developers that spans across industries and academic disciplines, the proposal is also purpose- and user-agnostic, permitting all users to perform machine learning techniques irrespective of whether they are associated with an academic institution or a commercial enterprise.

Clearing the Path for Hong Kong's AI Ambitions

By recommending the adoption of an express exception in the Copyright Ordinance to cover copying of a lawfully accessed work for the purpose of "computational analysis," CEDB can help enhance the competitiveness of Hong Kong's AI industry. Consistent with international best practice, CEDB should introduce an exception that is technologically neutral, sufficiently flexible so as to be future-proof, and agnostic as to purpose and user. With that in mind, the exception should extend to:

- Commercial and non-commercial uses;
- All works and other subject matter; all copyright-relevant acts, including retention of data for purposes of verification and validation of results; and,
- The provision of AI services (i.e., permitting service providers to perform computational analysis on behalf of end-users).

Importantly, such an exception would be consistent with Hong Kong's international obligations. The TRIPs Agreement and Berne Convention require Member States to ensure that exceptions to copyright are confined to "certain special cases which do not conflict with the normal exploitation of the work and do not unreasonably prejudice the legitimate interests of the rights holder."[20] An exception to facilitate computational analysis of lawfully accessed works is consistent with each of these requirements. It would meet the "certain special cases" requirement insofar as it is clearly and narrowly tailored to advance a significant public interest in the development of AI. It will not conflict with the normal exploitation of copyrighted works because reproductions made during an computational analysis process are not visible to humans and cannot substitute for or displace markets for

---

[19] Division 8, Section 233(e).

[20] TRIPs Article 13.

the original works. And finally, an computational analysis exception will not prejudice the legitimate interests of a rights holder because the value derived from such processes is unrelated to the expressive content that copyright is intended to protect. Indeed, the very purpose of performing computational analysis is identify *non-copyrightable information* – such as the relationships and correlations between large corpuses of works – that can be used to train AI models. The value derived from such activity lies not in the factual information that is gleaned from any single source, but rather in the discovery of entirely new forms of knowledge that emerge from the identification of patterns and correlations that exist between large bodies of disparate sets of data. While copyright protects the specific expression of factual information, it does not extend to facts themselves, and it was never intended to prevent users from analyzing a work to which they have lawful access in order to derive factual, non-copyrightable information. Once lawful access to a work is obtained, it should not matter whether a user analyzes the material manually or extracts the underlying factual information through a digital process.

<center>* * * *</center>

BSA appreciates the opportunity to provide feedback on these critically important issues and we look forward to remaining engaged as CEDB considers next steps.

Sincerely,

Christian Troncoso
Senior Director, Policy