Coalition for the Diversity of Cultural Expressions

Brief of the Coalition for the Diversity of Cultural Expressions
Consultation on Copyright in the Age of Generative Artificial Intelligence

A. CDCE

The Coalition for the Diversity of Cultural Expressions (CDCE) brings together the main English and French professional organizations in the cultural sector in Canada. It is composed of 54 organizations that collectively represent the interests of more than 360,000 professionals and 2,900 organizations and businesses in the book, film, television, new media, music, performing arts and visual arts sectors. The CDCE's main objective is to ensure that cultural goods and services are excluded from trade negotiations and that the diversity of cultural expressions is present in the digital environment.

The Coalition ensures that Canada retains the sovereign right to develop, implement and modify the policies, programs and measures required to ensure we have a robust supply of Canadian artistic expressions of all kinds, in every medium, and from all communities. CDCE also works to protect and promote our artists and cultural industries, and to ensure there is a rich diversity of cultural expressions in Canada and globally, including in the digital environment.

The *Copyright Act* is one of the key tools available to the Canadian government to foster a viable, sustainable, and diverse cultural ecosystem. The rapid developments in generative artificial intelligence over the past year will undoubtedly impact the diversity of cultural expressions in Canada. It is timely to question the robustness of the *Copyright Act* in this context. The CDCE, however, wishes to emphasize from the outset that this is not the only legislative tool that can or should be mobilized to protect the diversity of cultural expressions in response to these developments. The cultural sector requests to be included in all Canadian reflections surrounding the governance of AI.

Having stated these premises, the CDCE is pleased to respond to the Government of Canada's Consultation on Copyright in the Age of Generative Artificial Intelligence (the "**Consultation**").

B. Fact Finding: Recent Applications of Generative Artificial Intelligence (AI) in the Cultural Sector

The CDCE has been examining the interaction between AI and culture for several years. [See for example, CDCE (2018), Ethical Principles for the Development of Artificial Intelligence Based on the Diversity of Cultural Expressions.] Our members recognize AI's potential and are exploring

the benefits of this new technology. AI, like other tools, can be used to enhance and support human creativity when used responsibly and ethically. In the cultural industries, AI is used as a tool that supports – not replaces – the original human expression of creators' works. Some creators are using AI as a tool as part of their creative process to reduce some of the time-intensive, rote aspects of their work. Publishers and producers are using it to assist with layout, style, visual effects, color-correction, and sharpening detail, among other things. When used responsibly, AI can provide tremendous value to the creative process. However, when used irresponsibly, AI has the potential to seriously undermine and damage the cultural sector and the diversity of cultural expressions in Canada, and around the world.

This Consultation is focused on "**Generative AI**", that is, systems that use deep-learning models that can generate high-quality creative content based on the works, performances, and sound recordings they have been built on. In comparison, "traditional AI", generally refers to tools or systems that perform specific tasks based on predefined rules and inputs. Using AI to assist in crafting a response to an email is markedly different than using it to produce a song, an illustration, or a poem. When considering whether change is needed to copyright policy, there is a need for careful differentiation between the two. We are pleased that the Government has focused this Consultation on Generative AI and concentrate our response on that next generation of AI.

As a general principle, AI developments can and should co-exist with a copyright system that incentivizes creators to create and disseminate their work and protects the rights of copyright owners. But Generative AI platforms are profiting significantly from the unauthorized uses and reproductions of the works, sound recordings and performances represented by CDCE members.

As noted in the Consultation Paper, the text, and data mining (TDM) used for Generative AI systems involves the reproduction of large quantities of data and copyright-protected works, performances and sound recordings. These "inputs" in TDM are often copyright-protected works, or other subject matter, for which no licenses are sought, nor any compensation flowing to rightsholders.

TDM also infringes authors' moral rights. No attribution is given to the authors of the text, music, artwork, and other copyright-protected content that is ingested into Generative AI systems, either during the training of the system or during the use of the system. TDM also distorts works, including by cropping photographs, using lower resolutions, and disaggregating lyrics, text, or lines of music into segments and reassembling them into different sequences. These material alterations and mutilations offend the integrity of authors' works to the prejudice of their honour and reputation.

The outputs of Generative AI similarly pose fundamental and existential issues for the cultural sector. As examples, in the music sector, Jukebox, released by OpenAI in Beta form, can produce a "wide range of music and singing styles, and generalizes to lyrics co-written by a language model and an OpenAI researcher" precisely because it has ingested vast amounts of previously

composed and recorded music. [See Jukebox (openai.com)] Similarly, MuseNet can generate up to "4-minute musical compositions and can combine styles from country to Mozart to the Beatles". OpenAI states that MuseNet discovered patterns of harmony, rhythm, and style using the same technology as GPT-2, a large language model that ingested a dataset of 8 million webpages. [See: MuseNet (openai.com)]. Books3, used to train Meta's Generative AI, was based on a collection of more than 191,000 pirated books. [See: These 183,000 Books Are Fueling the Biggest Fight in Publishing and Tech - The Atlantic].

Generative AI outputs also often infringe creators' copyrights by reproducing and communicating to the public "lookalikes" and "soundalikes" of creators' original expression, often in response to prompts of users seeking exactly that.

In addition, various applications rely on AI, including TDM, to generate synthetic media, such as deepfakes, holograms, digital replicas, stand-ins, voiceovers, virtual characters, and environments. Applications can reproduce a person's voice, image and/or likeness. Text-to-speech systems can reproduce a person's voice, while other systems can rework the facial expressions of actors to assist with dubbing. If the output incorporates performer's image or likeness, it is a reproduction of a substantial part of that performer's performance and an infringement.

Finally, AI-generated output can also infringe authors' moral rights. As one example, if a work is marketed as "in the style of" or "in the sound of" a particular author, the use will often be a bastardization and inferior copy of the author's work. Infringing output can also be used in inappropriate contexts, like in a political campaign with which the author does not agree. Both examples could cause real prejudice to the honour and reputation of the author.

These types of outputs compete directly with the market for creators' works, and threaten the livelihoods of Canada's writers, authors, actors, publishers, musicians, songwriters, composers, visual artists, performers, directors, labels, music publishers, and producers.

All these uses, and the consequent infringements, must be considered when examining the copyright policy implications of Generative AI.

C. Proposed New Exceptions are Unwarranted and Unnecessary

The Consultation paper asks, "What would more clarity around copyright and TDM in Canada mean for the AI industry and the creative industry?" Absolute clarity around copyright and TDM in Canada need not be the goal of this Consultation, or of copyright policy more generally, especially as the market is developing around these new uses. Serious risks arise from assuming that a new form of technology automatically requires the creation of a new exception. Reflexive approaches do not, and cannot, consider the speed at which Generative AI technology is evolving and the resulting impact on affected markets.

Moreover, introducing a new exception for TDM now would interfere with the ability of market participants, namely users and rightsholders, to set the boundaries of that emerging market. It would be particularly disruptive for the Government to introduce a new exception for TDM as licensing models are developing and emerging. Various licensing models for the use of copyright-protected works, performances and sound recordings in Generative AI already exist in the market, including:

- the Copyright Clearance Center (CCC) in the United States has offered TDM licences for the use of literary and artistic works for several years;
- many Scientific, Technical and Medical journal publishers have been offering TDM licences in Canada for a long time. [See for example, Elsevier's TDM licence and Taylor & Francis's TDM licence];
- Getty Images has made a Generative AI platform that uses only its own licensed images [Getty Images Launches Commercially Safe Generative AI Offering - Getty Images]
- Meta's MusicGen reportedly used Meta-owned and licensed music from stock footage and music licensing companies [Introducing AudioCraft: A Generative AI Tool For Audio and Music | Meta (fb.com)]; and
- Universal Music Group has recently announced collaborations with AI developers to explore how their technology can promote and enhance the creative process, including a partnership with Bandlab, a social music making platform, with a mandate to "advance the companies' shared commitment to ethical use of AI and the protection of artist and songwriter rights". [See here: UNIVERSAL MUSIC GROUP AND BANDLAB TECHNOLOGIES ANNOUNCE FIRST-OF-ITS-KIND STRATEGIC AI COLLABORATION - UMG].

Given the nascent market for the licensing of TDM activity and Generative AI partnerships that are developing between AI developers and rightsholders, now is most certainly not the time for the Government to step in and introduce new exceptions. Instead, the Government should allow the market to work out market-based licensing solutions for TDM uses in Generative AI.

According to the Consultation Paper, "Because of the large quantity of data often involved in training such models, in particular when sources from the Internet, obtaining any necessary authorization from rights holders to make reproductions of the works or other subject matter in the course of these activities could be a significant burden". Given the above examples, this is most certainly not the case.

Rather, it appears that those who are calling for exceptions prefer to lobby governments around the world as opposed to coming up with market-based solutions. After all, sometimes market-based solutions come with a price. Exceptions are usually free.

The Consultation Paper raises the 2019 EU Directive that requires member states to provide two TDM exceptions: one for research and heritage institutions, and one for any other person and

purposes, which rightsholders can "opt out" of. Any suggestion of an "opt-out" regime for TDM is both controversial and impractical. The introduction of an exception that gives rightsholders the ability to "opt-out" of such an exception turns copyright on its head. Copyright is an opt-in system: no formality is required for a work to attract copyright protection. Requiring a copyright owner to advise a platform that it objects to the use of its work in TDM for Generative AI is a formality that violates Canada's obligations under the Berne Convention. There is no basis to throw out these fundamental principles of copyright law and Canada's international treaty obligations.

An opt-out system would require all copyright owners to monitor every Generative AI platform available in Canada and send some sort of notice to each Generative AI operator or application developer advising that it has chosen to "opt-out" of the exception for TDM purposes. As discussed below, copyright owners would first have to know that their works, performances, or sound recordings were being used by the Generative AI operator/application developer, or otherwise send a notice to all of them. Rightsholders would also have no remedy regarding any copying that took place before they opted-out. That is a massive burden to put on copyright owners and is wildly disproportionate to this supposed problem.

There have been suggestions in other jurisdictions that a compulsory licensing regime may be appropriate for TDM. Compulsory licensing deprives creators and copyright owners of their exclusive rights to authorize the reproduction of their works, performances, or sound recordings by forcing them to license, deprive them of fair compensation for those uses. It similarly deprives rightsholders of the ability to prohibit the use of their content by a service that might ultimately be cannibalizing the need for their own labour or that produces content that acts as a substitute for their original work. The implementation of such a regime would also impose a significant burden on copyright owners to administer and enforce. Additionally, implementing compulsory licensing for TDM is a solution in search of a problem. Compulsory licensing might make sense in certain special cases when voluntary licensing is impossible, such as the use of copyright-protected works or other subject-matter by retransmitters contemplated under section 31 of the *Copyright Act*. But here, licensing is not impossible: there is a functioning and growing market for licensing for TDM uses.

There is no reason to conclude that the current law is insufficient to address any uses that might arise with respect to the training of Generative AI. The *Copyright Act* is sufficiently technologically neutral to accommodate technological development and foster Generative AI innovation. As noted in the Consultation paper, there are existing exceptions in the *Copyright Act* that may assist users in appropriate cases. Until and unless a Canadian court or the Copyright Board exposes a real deficiency with respect to Generative AI that needs to be addressed, there is no valid policy reason to introduce any exception for TDM.

As AI developers and platforms are keenly aware, entering any market involves assuming responsibility for the impact of their new technology on that market and the players within it. The platforms must be responsible for respecting the copyrights of creators.

Numerous exceptions in Canada's copyright laws have already caused a structural imbalance in the *Copyright Act*: an imbalance that is depriving rightsholders of their "just reward" and that may otherwise discourage and disincentivize copyright owners from creating and disseminating their works, performances, and sound recordings. The introduction of any additional exceptions for TDM purposes would only serve to further upend the balance sought in the *Copyright Act*.

There is no need to introduce exceptions allowing further uses of rightsholders' works, performances, or sound recordings in Generative AI systems.

Recommendation 1. That the Government neither amend the current exceptions to include TDM, nor implement new exceptions for TDM.

D. Questions of Fair Compensation Should Be Left for the Market to Determine

With respect, when and how rightsholders should be compensated for the use of their works, performances, or sound recordings as inputs into Generative AI systems is not a question that should be answered by the Government. Nor should the Government interfere with the question about what level of remuneration is appropriate. The compensation and fair remuneration required for the use of a given work, sound recording, or performance will, and should, be set by the market or, if need be, the Copyright Board of Canada.

This question is properly the subject of negotiations between rightsholders and Generative AI platforms. There may indeed be instances where rightsholders agree with a platform that compensation is not necessary. But the Government should not decide that question. Doing so would simply impede those negotiations and prevent a market-based solution for TDM.

But additionally, it is not just compensation that should be the focus of the question. Among other things, the *Copyright Act* provides rightsholders with exclusive rights to reproduce their works, performances and sound recordings, or any substantial part thereof, and to authorize any such acts. These rights are engaged when copyright-protected works, performances and sound recordings are ingested into Generative AI systems. Authorization and permission are as important as compensation, particularly when the system output might compete with, or act as a substitute for, the original work. Again, market negotiations and the developing licensing market ought to decide these questions, not the Government.

E. Challenges with Licensing, Monitoring and Enforcement

The Consultation Paper asks whether rightsholders are facing challenges in licensing their works or related rights for TDM activities.

Many platforms appear to believe the ingestion of copyright-protected content in their systems is already exempted, or that it does not require authorization, under copyright law in Canada. Globally, Google is so confident that the inputs to some of its Generative AI platforms and the output generated by them are non-infringing that it is indemnifying its users from copyright infringement claims:

> "If you are challenged on copyright grounds, we will assume responsibility for the potential legal risks involved. To do this we will employ a two-pronged, industry-first approach designed to give you more peace of mind when using our generative AI products. The first prong relates to Google's use of training data, while the second specifically covers generated output of foundation models. Taken together, these indemnities provide comprehensive coverage for our customers who may be justifiably concerned about the risks associated with this exciting new frontier of generative AI products." [See: Google]

Microsoft is similarly confident, offering to "defend its customers from intellectual property infringement claims arising from the customer's use and distribution of the output of its Generative AI Copilot services". [See: Microsoft].

Notably, the former VP of audio at Stability AI (the maker of the popular image generator Stable Diffusion) announced that he recently resigned from his role at Stability AI because he did not agree with the company's opinion that ingesting copyright-protected content into Generative AI models is fair use. [See Stability AI VP quits in 'fair use' copyright protest • The Register]

These types of ingrained and bullish positions pose challenges for rightsholders in licensing their works, performances, and sound recordings for TDM.

In addition, rightsholders have no insight into whether their works, performances or sound recordings have been ingested into any Generative AI platform. TDM on any platform is a black box. This information asymmetry, and resulting imbalance in negotiating power, makes monitoring and the resulting licensing opportunities incredibly difficult for rightsholders. At best, licensing transactions are inefficient: rightsholders are left with the choice of guessing whether a particular Generative AI system has used their works or waiting for the operator of a system to approach them for a licence. At worst, there is complete market failure where operators are free riding off the backs of creators' content. This imbalance must be corrected. For all these reasons, we recommend the Government apply legally binding transparency obligations to Generative AI systems, like those that were recommended by the European Parliament and

contained in the provisional agreement for foundation models [See: European Parliament and European Council]

Recommendation 2. That Generative AI platforms be required to comply with transparency requirements, including, but not limited to: (i) publishing records of the copyright-protected works, sound recordings and performances that were ingested into the platform; (ii) designing the model to prevent it from generating illegal or infringing content; and (iii) disclosing that the output produced by the system was generated by AI.

These obligations ought not be limited to "high-impact systems" as defined and contemplated in the *Artificial Intelligence and Data Act* (AIDA) proposed in Bill C-27 but must apply to all Generative AI large language models.

The imposition of transparency requirements should not cause any real challenges for developers because they already document this data. As one example, the GPT-2AI model card published by Open AI in 2019 includes a list of the top one thousand domains present in their dataset, as well as their frequency. In this log, you will find illegal sites (Pirate Bay), pornography (YouPorn) and sites of rightsholders (Le Monde, CBC) [See Github] By requiring transparency, not only will rightsholders have access to essential information for the management of their copyrights and related rights, but users of the systems will have critical information about the sources and biases that may be inherent in the system itself.

While these proposed transparency requirements are not all copyright-specific, and may instead more properly be the subject of Bill C-27 or other legislation, increased transparency in the inputs into and use of Generative AI will help ensure the systems are responsible, lawful, safe, transparent, accountable, and non-discriminatory.

F. Attribution/Ownership/Authorship

The questions asked in the Consultation regarding authorship and ownership of copyright in AI-generated content raise fundamental questions for the cultural sector.

While the *Copyright Act* does not explicitly define the word "author", Canadian case law has already reiterated that original works protected by copyright must be the product of an author's exercise of skill and judgment, which "must not be so trivial that it could be characterized as a purely mechanical exercise". [*CCH v. The Law Society of Upper Canada*, [2004] 1. S.C.R. at para. 25.]

Debates on this issue are ongoing at WIPO and in many other jurisdictions, including the United States and the European Union. However, the international consensus is that human authorship is core to copyright and that content generated by AI without any human involvement is not, and should not be, protected by copyright. CDCE agrees with this consensus. The same principles apply to performer's performances: only human performances are entitled to rights and protections under the *Copyright Act*.

The purpose of copyright is, in part, to obtain a just reward for the creator (and prevent someone other than the creator from appropriating whatever benefits may be generated) and to incentivize further creation. At this point in time, there is no need to amend the *Copyright Act* to create new rights to incentivize the creation of AI-generated content.

Granting copyright or related rights to Generative AI systems for autonomous or mechanical content, without original expression of an idea attributed to a human author or performer, would shift the copyright regime from a paradigm of protecting and promoting human creativity to the pursuit of innovation and revenues for companies of all kinds. This would have far-reaching consequences, the long-term impact of which is difficult to anticipate.

Finally, there is something perverse and frankly, offensive about suggesting that there should be an exception to human authors' and performers' copyrights and related rights for the inputs into Generative AI systems, while also providing the same platforms with additional copyright protections for non-human generated outputs. Leaving aside the resulting job losses in the sector, this prospect of mass production of pseudo-cultural content entirely generated by Generative AI systems is a major social concern. "Creation" would be the result of companies seeking solely to make their products mass marketed and profitable, as opposed to having a multitude of diverse creators and artists expressing their own thoughts, views, opinions, commentary, and creativity.

Products resulting from purely mechanical AI generation processes that lack original human expression are not "works" protected by copyright or any sort of neighbouring right and should not become so.

Recommendation 3. That the Government not amend the Act to afford copyright protection to AI-generated content.

In addition, it is important that performer's performances continue to have rights and protections like those they current enjoy under the *Copyright Act*, whether or not the underlying content is generated by AI.

Recommendation 4. That performers' performances remain fully protected under the *Copyright Act*, including when the content performed is AI-generated.

G. Infringement and Liability.

To establish infringement, a rightsholder must prove that the defendant copied or made available all or a substantial part of a work, performance, or sound recording, that the defendant had access to the original work, performance or sound recording and that the original work,

performance, or sound recording was the source of the copy. Independent creation is a complete defence to infringement.

The biggest barrier to determining whether an AI system accessed or copied a specific copyright-protected work, sound recording, or performance is the lack of transparency described above. Without some knowledge of the copyright-protected content ingested into a Generative AI system, rightsholders can only suspect that their content has been used without authorization. In some cases, this will lead to undetected or unproveable large-scale infringements. It will also lead to a process where rightsholders are forced to sue Generative AI platforms they assume have infringed their copyrights in to (hopefully) obtain disclosure in discoveries, incurring significant time, resources, legal and expert fees, and expenses. At the end of that process, a rightsholder may find out the platform never had access to the copyright-protected content in the first place. This would result in a wildly inefficient system and cannot be the intent of the Government.

Requiring Generative AI systems to publish records of the copyright-protected content that was ingested into the systems is necessary so that copyright owners can protect and monetize their intellectual property. With these transparency obligations in place, liability could potentially arise for primary, secondary, or authorizing infringement, enabling infringement, moral rights infringement, removal of digital rights management information, and circumvention of technological protection measures. From an infringement perspective, the current *Copyright Act* will be sufficient to address issues specific to Generative AI provided these record-keeping and disclosure obligations are in place.

Perhaps more importantly, transparency requirements will also promote a functioning and more efficient licensing market where rightsholders and users can negotiate on a more level playing field by reducing the information asymmetry that is currently present in the market.

Finally, the Government must consider the impact of Generative AI on authors' and performers' moral rights, including their rights to the integrity of their works and performances, as well as name and likeness rights and rights of personality and publicity. While not nearly a full answer to these issues, transparency obligations and disclosure requirements will at least signal that uses like deepfakes and voiceovers, are not sanctioned by the performers they are imitating.

Ultimately, the aim of this Consultation on Generative AI should be to encourage responsible use of copyright protected works and a healthy, functioning licensing market for the use of copyright-protected works, sound recordings and performances in TDM. Transparency obligations that require Generative AI platforms to disclose records of the copyright-protected content that was ingested – as opposed to the creation of new exceptions – is the best way to ensure that Generative AI can continue to innovate alongside a copyright system that incentivizes creators to create and disseminate their work and that provides them with their just reward.

H. Other *Copyright Act* Amendments

The CDCE has made other recommendations to improve the *Copyright Act*, which are reproduced in Appendix 1. We request that the next copyright reform include these other recommendations, even if they are not subject to the current consultation.

We thank the Government for this opportunity to provide our comments on this important Consultation.

Appendix 1

The six urgent recommendations of CDCE members:

1. Amend the fair dealing provisions in the context of education so that they only apply where a work is not commercially available under a license by the rightsholder or a collective society.

2. Incorporate resale right into the *Copyright Act*.

3. Abolish the public performance royalty exemption for performers and producers for commercial radio stations.

4. Amend the definition of sound recording to include sound recordings that accompany audiovisual works.

5. Amend the Act to confirm the binding nature of tariffs set by the Copyright Board.

6. Restore the private copy regime in the music sector.

The six mid-term recommendations of CDCE members:

1. Ratify the Beijing Treaty and grants moral and economic rights to performing artists on audiovisual media in the Act.

2. Raise the upper and lower limits of statutory damages for non-commercial violations and allow the establishment of higher damages in case of systematic and massive use.

3. Ensure that right holders in the various sectors have the same tools by ensuring that all collecting societies can claim statutory damages of three to ten times the value of the tariff that has not been paid.

4. Improve the private copying regime by allowing the payment of royalties for rights holders in the audiovisual, literary, and visual arts sectors.

5. Amend the exemption in section 32.2(3) to limit its application to acts without motive of gain.

6. Take into account the needs and realities of Indigenous artists, creators, and organizations.