



Canadian Government Consultation on AI Response to On-line survey

Technical Evidence Questions:

- How does your organization access and collect copyright-protected content, and encode it in training datasets?
 - How does your organization use training datasets to develop AI systems?
 - In your area of knowledge or organization, what measures are taken to mitigate liability risks regarding AI-generated content infringing existing copyright-protected works?
 - In your area of knowledge or organization, what is the involvement of humans in the development of AI systems?
 - How do businesses and consumers use AI systems and AI-assisted and AI-generated content in your area of knowledge, work, or organization?
-
- *At this point in their development, AI applications are being developed by CPC members for their potential contribution to greater efficiencies in the publishing business, across Trade, Educational and Professional parts of our sector. Examples are provided below. As a critical principle, CPC members approach this with complete adherence to existing copyright law and a commitment to preservation of creators' exclusive rights, and argue for the paramount importance of these principles in this submission.*
 - *Areas of concern include the unlicensed use of works for training AI systems. In the educational field this includes the use of AI systems to circumvent Assessment tools in areas such as exams and assignments across the whole education spectrum but particularly in Post-Secondary Education. This is a concern shared of course by Post-Secondary institutions—universities and colleges, both public and private—who are important customers for many CPC member firms.*
 - *CPC members are using their own content to train AI systems or when more training data is required, will enter into licenses to be able to legally use copyright content. We note that the developers of the large language models (known as LLMs and including OpenAI, Meta and Google) unfairly compete with publishers by using infringing copies of books and other media such as books that are being illegally made available through pirate online libraries like Books2 and Books3.*

TDM Questions:

- What would more clarity around copyright and TDM in Canada mean for the AI industry and the creative industry?

Our principal concern will always be the upholding of rightsholders’ rights—both authors/creators and publishers—as the market for AI-generated content evolves. It is critical that rightsholders’ consent be obtained for training AI systems. In our view, training AI systems involves multiple reproductions and adaptations of works. There is no question that copying is involved when works are downloaded for training purposes, when works are copied and compressed into AI models, and in many cases the outputs from systems training on publishers’ materials are reproduced verbatim or are adaptations that include substantial parts of ingested materials because the AI systems’ “memorization” capabilities.¹

Further, in our view, the models training, the models that are trained using publishers’ works, and many of the outputs would infringe copyright, subject to any defenses such as fair dealing for the purpose of research (discussed below). It must be borne in mind that a work is reproduced when it is copied in “any material form.” This means that a work is reproduced even when it is encoded in any language or notation (including tokens generated during AI training) or is stored in a compressed form from which publishers’ works can be perceived by or with the aid of a machine or device.² Thus, training AI models using unlicensed publishers’ works implicates the reproduction right because of the three different processes involved in training AI models and providing services using them that generate outputs. Further, operators of generative AI systems that produce synthetic outputs that contain reproductions of all or any substantial parts of publishers’ works may also infringe the authorization right and/or the making available rights under the Copyright Act.³

With that said, any litigation necessary to vindicate publishers’ exclusive rights would be costly and would likely take many years from when a case is commenced until it is ultimately resolved by the Supreme Court. In the interim, many AI operators including the very large platforms will likely continue to exploit publishers’ works without consent or providing any form of compensation. Further, publishers would be unable even to collectively license their works for training purposes

¹ See, [Exhibit J to the Complaint by the New York Times against OpenAI and Microsoft](#); also, [Extracting Training Data from ChatGPT](#).

² The *Copyright Act* makes it an infringement to reproduce a work (or a substantial part of a work) in any “material form”. This is a technologically and media neutral term that would make any copying into a form of storage a reproduction, regardless of the form of notation or encoding. See, *Apple Computer Inc. v. Mackintosh Computers Ltd.* (1986), 28 D.L.R. (4th) 178 (Fed. T.D.), varied (1987), 44 D.L.R. (4th) 74 (Fed. C.A.), affirmed [1990] 2 S.C.R. 209; *Canadian Broadcasting Corp. v. SODRAC 2003 Inc.*, 2015 SCC 57, *Robertson v. Thomson Corp.*, [2006] 2 S.C.R. 363; *Labrecque (O Sauna) c. Trudel (Centre Bellaza, s.e.n.c.)*, 2014 QCCQ 2595.

³ If the outputs are copied by users, then the AI operators infringe the authorization right. If copies are only made available for viewing, then the AI operators may be liable under the making available right. See, [Society of Composers, Authors and Music Publishers of Canada v. Entertainment Software Association](#), 2022 SCC 30.

using the Copyright Act's tariff system because of the unfortunate decision by the Supreme Court that held that even certified tariffs are not mandatory.⁴ While the law is clarified by the courts, AI operators may continue to unfairly copy publishers' works and thereby compete with publishers' commercialisation of their publications and undermine their own adoption of AI systems. We note that France's National Assembly Bill aimed at regulating artificial intelligence by copyright⁵ would expressly require the authorization of authors or rights holders for training or exploitation of artificial intelligence systems.

For the above reasons, the CPC recommends that the law be clarified to provide an express right of publishers to authorize the use of their works in AI systems that are trained or made available for commercial purposes ("commercial AI use"). Further, to avoid providing foreign AI operators an advantage over Canadian operators, the training authorization right should apply to any AI systems that make the service available in Canada.

As noted above, some uses of publishers' works could fall within a fair dealing exception. The Supreme Court has construed the fair dealing exception for research and the fair dealing factors as user rights. Under the Supreme Court precedents, the exception for research may include activities that are not limited to non-commercial or private contexts and can include those that "facilitate" research and even large-scale infringements can, in certain instances, also be considered to be fair.⁶ However, publishers view the research exception as intending to promote the goal of research conducted by human beings and do not accept that training machines even to facilitate research by human beings is compatible with the human centric goal of research under the fair dealing exception. However, the law on this has not yet been settled in Canada. **Of course, all fair dealing determinations are questions of fact and "a matter of impression" and are left to the courts. Therefore, publishers recommend that the clarified exclusive right to authorize commercial AI use not be subject to any fair dealing exception.**

The CPC strongly objects to any TDM exception. It would be impossible to properly calibrate such an exception to take into account all of the evolving AI business models and uses of publishers' works for AI purposes. Further, given the evolving market for publishers' own exploitation and licensing of their works for AI purposes, any TDM exception would undermine market forces and effectively create an uncompensated compulsory license. A TDM exception could violate Canada's TRIPs and other treaty obligation that prohibits any exceptions to copyright that would violate the Three Step Test.

⁴ *York University v. Canadian Copyright Licensing Agency (Access Copyright)*, 2021 SCC 32 ("York")

⁵ France National Assembly Bill No 1630 online @

https://www.assemblee-nationale.fr/dyn/16/textes/l16b1630_proposition-loi "Article L. 131-3 "The integration by artificial intelligence software of intellectual works protected by copyright into its system and a fortiori their exploitation is subject to the general provisions of this Code and therefore to the authorization of authors or rights holders".

⁶ *Society of Composers, Authors and Music Publishers of Canada v. Bell Canada*, 2012 SCC 36; *York*

It is also critical that companies developing AI systems be required to accurately track, retain and disclose content sources used for AI training, without any attempt to rely on existing exceptions (e.g., Research, Education) to circumvent this requirement. In the absence of this protection, it is impossible for rightsholders to know if their works have been ingested for training purposes, given the vast scope of inputs into Generative AI models, and the 'black box' nature of their modeling.

There are several ways to help promote transparency. First, there should be a Bill of Discovery right to enable rights holders to obtain disclosure as to whether their works were used to train AI models.

Second, amendments could be made to [Canada's draft Artificial Intelligence Act \(AIDA\)](#) to expressly include transparency requirements on persons who make available or manage a general-purpose system along with the other transparency requirement currently being proposed by the Minister. [AIDA should also require such persons to respect Canadian copyright laws regardless of where the training and deployment of the AI systems takes place if the AI system is deployed in Canada.](#) [Publishers also view the unlicensed use of their copyright content to be very damaging and recommend that the term "harm" in AIDA expressly include the economic and moral harms to authors and publishers associated with generative AI systems.](#) It is only in these ways that rightsholders can be assured of the opportunity to participate in the value chain created by use of their content for Generative AI.

- Are TDM activities being conducted in Canada? Why or why not?

TDM activities are undoubtedly being conducted in Canada, in a broad range of applications that include media monitoring and scholarly research, especially in as Large Learning Models. Some of our members are developing their own AI solutions. For example, Thomson Reuters offers services called "Westlaw Precision" and "CoCounsel" which jumpstarts AI assisted legal research. A second offering built on Open AI's GPT-4 provides legal professionals with document review, legal research, contract analysis, compliance and other functionality. Most of our members would also engage with existing AI platforms such as OpenAI's ChatGTP.

- Are rights holders facing challenges in licensing their works for TDM activities? If so, what is the nature and extent of those challenges?

The principal challenge here is ensuring that exclusive rights under copyright law are respected. This includes tracking and disclosure of content use for ingestion into AI models, as mentioned above. Many providers of generative AI systems disregard copyright rights in the hopes they can become well established and eventually forestall having to pay for their unlicensed uses of works. The sooner

the law is clarified, the easier it will be for publishers to license their works including for uses in Canada.

- What kind of copyright licenses for TDM activities are available, and do these licenses meet the needs of those conducting TDM activities?

Existing licensing mechanisms, both direct and collective, are available today to cover content licensing needs for AI model ingestion. For example, the Copyright Clearance Center (CCC) offers TDM licences for the use of literary works. Many scientific, technical, and medical publishers (STM publishers) offer TDM licences in Canada. An example is Elsevier’s TDM licence which permits non-commercial research.⁷ Taylor & Francis offers a TDM license which can also include a license for commercial purposes.

Further there are market developments that illustrate that developers of generative AI systems can develop AI systems using their own content or content that has been properly licensed from others. Examples are Getty Images (which uses its extensive library of licensed images),⁸ Adobe (with its Firefly generative AI model that is trained on Adobe Stock images, openly licensed content, and public domain content),⁹ Meta’s AI image generator (which was trained on public facing Facebook and Instagram images),¹⁰ and, as noted above, Thomson Reuters services called “Westlaw Precision” and “CoCounsel” (which are trained on its own proprietary content and public domain content it makes available).

We believe it is critical that market forces be allowed to develop and to facilitate appropriate licencing models for AI content ingestion, based on the framework of exclusive rights provided by copyright law. It must be remembered that a TDM exception will provide no remuneration or compensation to authors or publishers. The copyright law provides a market framework by which market prices are set for the uses of copyright works. A TDM exception would operate extremely unfairly by undermining publishers’ own ability to exploit their works including for AI purposes. It would also, effectively, operate as a very unfair wealth transfer mechanism from authors and publishers to AI companies, many of which are very large and well-funded companies. If these or any other companies want to use AI works for training, they should negotiate fair agreements and pay publishers for inputs the same as they do for other inputs.

⁷ <https://www.elsevier.com/about/policies-and-standards/text-and-data-mining/license>

⁸ [Getty made an AI generator that only trained on its licensed images - The Verge](#)

⁹ [Adobe Firefly - Free Generative AI for creatives](#)

¹⁰ [Meta’s new AI image generator was trained on 1.1 billion Instagram and Facebook photos | Ars Technica](#)

- If the Government were to amend the Act to clarify the scope of permissible TDM activities, what should be its scope and safeguards? What would be the expected impact of such an exception on your industry and activities?

Clearly content creators such as authors and publishers would want TDM activities to be conducted under licence only, and not under any newly conceived exceptions. The expected impact is clear and consistent with existing law: the protection of rightsholders' exclusive rights and the concomitant opportunity to ensure proper compensation for use of their content. These rights and their safeguarding are the foundation that supports the creative industries broadly, and publishing in particular, and is the legal framework that encourages creative endeavours which in turn are critical to Canadian culture and voices. Moreover, this approach reinforces the notion that human creativity is at the heart of what Copyright has been established to protect, and will continue to be the critical input for the cultural sector even as Generative AI exerts more influence in the market for content. Lastly, it is only when AI Developers engage in appropriate licensing mechanisms that fulsome investment in AI development will continue, and consumer confidence in AI outputs will be strengthened.

However, notwithstanding the foregoing, any TDM exception should not apply unless (i) the exception is purely for non-commercial research, (ii) the work used is not infringing and is not in breach of any license or contractual prohibition, (iii) any copying must be ephemeral, (iv) the resulting output cannot reproduce all or any part of the work used for training, (v) the copyright holder must be notified before the work is used, and (vi) there is no circumvention of any technological protection measure (TPM) or violation of the rights management information (RMI) provisions of the Act.

We also note that Canada is obliged by international treaties (including the WCT and WPPT and the CAUSMA) to provide the latter protections, and any TDM exception must respect these obligations as well as the Three Step Test. Further, we recommend that if a TDM exception is to be proposed it should be subject to these conditions:

- *Before a work or works from a publisher is used, the publisher must be notified and the AI company must have an obligation to negotiate a license at fair market value. If a price cannot be agreed, the publisher can have the price settled by arbitration or the Copyright Board. We note that copyright owners of books made available in Canada are listed in publicly available sources such as Booknet.*
- *Neither the TDM exception (nor any fair dealing exception) should apply if the work is commercially available under a license including a tariff approved by the Copyright Board.¹¹*
- Should there be any obligations on AI developers to keep records of or disclose what copyright-protected content was used in the training of AI systems?

¹¹ There is a precedent for this. See ss. 29.4(3), 30.1(2P) of the *Copyright Act*.

This requirement is absolutely critical to the evolution of legitimate AI industry in Canada and beyond. Clear and discoverable records of content use, with full disclosure requirements, must be the foundation for this framework in order to protect rightsholders' rights and ensure their ability to participate in the value chain that AI is developing. In the absence of such safeguards, human-based creativity is put at risk and may become increasingly marginalized.

- What level of remuneration would be appropriate for the use of a given work in TDM activities?

Market-based licensing mechanisms must be allowed to develop in this sphere. If the playing field is balanced between rightsholders and AI systems via the markets created by the system of exclusive rights, the price will set itself via negotiations. The appropriate price should always be the price that is (or would be) negotiated by a willing buyer and a willing seller.

Authorship and Ownership of Works Generated by AI Questions

- Is the uncertainty surrounding authorship or ownership of AI-assisted and AI-generated works and other subject matter impacting the development and adoption of AI technologies? If so, how?

At this point in time, there is no evidence of an adverse impact on the development or adoption of AI technologies due to the copyright status of generated works. However, rights holders would be concerned about providing copyright protection for works created entirely by computer-generated content.

- Should the Government propose any clarification or modification of the copyright ownership and authorship regimes in light of AI-assisted or AI-generated works? If so, how?

We are of the view that the Supreme Court of Canada originality test is fit for purpose and would adequately address the copyright status of AI-assisted works. We believe that copyright should be reserved for human-created works that meet the test of originality. AI-generated works that have NO human intervention should not be protected by copyright.

- Are there approaches in other jurisdictions that could inform a Canadian consideration of this issue?

The approaches in the UK and other countries that provide protection for computer-generated works were conceived before generative AI was introduced and would potentially provide protection for works created without the requirement of originality, which would not be appropriate.

Infringement and Liability Questions

- Are there concerns about existing legal tests for demonstrating that an AI-generated work infringes copyright (e.g., AI-generated works including complete reproductions or a substantial part of the works that were used in TDM, licensed or otherwise)?
- What are the barriers to determining whether an AI system accessed or copied a specific copyright-protected content when generating an infringing output?

There are no compulsory requirements for GenAI engines to record and disclose the datasets that are ingested for AI training purposes. In the absence of this requirement, it is difficult to determine what copyright-protected content has in fact been used. This adversely affects the ability of publishers to negotiate licenses or seek legal redress. We recommend the approach being adopted in the EU under the draft Artificial Intelligence Act (AIA) that would require transparency, except to the extent a creator uses its own works for training purposes.¹²

- When commercialising AI applications, what measures are businesses taking to mitigate risks of liability for infringing AI-generated works?

¹²— “4. Providers of foundation models used in AI systems specifically intended to generate, with varying levels of autonomy, content such as complex text, images, audio, or video (“generative AI”) and providers who specialise a foundation model into a generative AI system, shall in addition:

c) without prejudice to national or Union legislation on copyright, document and make publicly available a sufficiently detailed summary of the use of training data protected under copyright law.” EU AI Act (draft Compromise Amendments) May 9, 2023

The CPC members are vigilant only to use their own works or works that are licensed. This is not the practice of the large Generative AI companies like OpenAI, Meta's LLM, or Google. There are instances of AI-generated content, derived from AI training from copyright-protected works, being fraudulently marketed as being authored by the creator of that copyrighted work (e.g., Clare Duffy lawsuit in U.S.).

- Should there be greater clarity on where liability lies when AI-generated works infringe existing copyright-protected works?

The extensive litigation in the United States shows that AI companies reject that outputs from their content infringe copyright. Some outputs clearly reproduce all or a substantial part of input data. The New York Times suit (referred to above) is a good example. However, grey area is where the outputs are derived from the training data, but specific outputs may not actually meet the test to be a reproduction under the Copyright Act. For example, "style" may not be protected by copyright. But generative AI systems are capable of and do produce output in the "styles" of authors, artists, and other creators. Further, "facts" are not protected by copyrights. But, here again, generative AI systems can wholly ingest facts as part of the ingestion of works and use these facts in output in ways that could compete with the publishers, a good example being a publication that is comprised of compilations. In a broader sense, all generative AI content is derived from copyright works or a mixture of copyright works and works in the public domain (such as where the copyrights have expired) or works that have been licensed.

This "derived" content is extremely valuable and it would be very unfair for the copyright works to be appropriated to generate massive profits for AI companies with no control or compensation to publishers and other creators. A major challenge with the existing copyright law is that while a specific output may not be derived and reproduce a specific input, when taken in the aggregate many outputs would be derived from the aggregation of unlicensed copying of publishers' and other creators' copyright works. It is undoubtedly for this reason that France's National Assembly Bill aimed at regulating artificial intelligence by copyright¹³ would provide for compensation to be paid to authors for all generative AI content whose origin cannot be determined.¹⁴

¹³ France National Assembly Bill No 1630 online @ https://www.assemblee-nationale.fr/dyn/16/textes/l16b1630_proposition-loi

¹⁴ Article L. 121-2 ""Moreover, in the event that a work of the mind is generated by an artificial intelligence device from works whose origin cannot be determined, taxation intended to promote the creation is established at the benefit of the body responsible for collective management designated by amended article L. 131-3 of this code. "This taxation is imposed on the company which operates the artificial intelligence system which made it possible to generate said "artificial work"

This issue is an especially important one for Canadian publishers as, to date, much of the training of LLMs has been done in the United States. This leaves Canadian-based publishers with few remedies in Canada for output that is largely derived from copyright works including where specific outputs are not infringing. Also, as generative AI systems will often return different results from specific text-based prompts, it may be difficult to make out a case of reproduction by relying solely on a specific output.

The publishers therefore recommend that the Copyright Act be clarified so that all outputs of generative AI systems that are derived from copyright works, individually or in the aggregate, used without a license be deemed to be infringing reproductions.