

Copyright and Generative AI Consultation

Questionnaire Responses from the Canadian Archival Community

January 14, 2024

Technical Evidence

The Government of Canada invites views on technical aspects of AI technologies, including on the following questions:

- **How does your organization access and collect copyright-protected content, and encode it in training datasets? NA (not applicable)**
- **How does your organization use training datasets to develop AI systems? NA**
- **In your area of knowledge or organization, what measures are taken to mitigate liability risks regarding AI-generated content infringing existing copyright-protected works? NA**
- **In your area of knowledge or organization, what is the involvement of humans in the development of AI systems? NA**
- **How do businesses and consumers use AI systems and AI-assisted and AI-generated content in your area of knowledge, work, or organization?**

The holdings in Libraries, Archives and Museums (LAMs) are a major source of documents for AI researchers in their development of training datasets for use in training AI models, particularly those datasets used to train large language models. Canadian archival holdings are a rich treasure trove of records in all formats that serve as the raw material for scholars, students, and ordinary citizens. Archival institutions have eagerly embraced the opportunities provided by the Internet to digitize and make our holdings available online. When digitized, traditional records can be mined for valuable historical information. For example, fur traders' journals document decades of weather patterns, relations between indigenous people and settlers, and early commercial activities; and tax rolls record the names of residents, which are of great interest to family historians. In addition, archival institutions are acquiring born-digital records and research data sets from their parent institutions and private donors.

Transparency and ensuring non-bias in training datasets is of concern to LAMs because of our public service mission. Although it may be onerous, it is very

desirable to create the metadata to be able to identify and link the training datasets to the generative AI output created by the AI tool. This metadata will provide transparency and oversight to the users, and also for the creators and rightsholders whose works are used in the datasets in a non-consumptive way.

Canadian LAMS are also beginning to use some of the emerging generative AI tools for their own purposes. Archives can use AI tools to generate basic metadata, transcriptions, and create important access points for all manner of digitized documents, thereby providing greatly improved access to our holdings for their researchers and the General Public. (See Pavis, Mathilde. Artificial Intelligence: a digital heritage leadership briefing, 2023 <https://www.heritagefund.org.uk/about/insight/research/artificial-intelligence-digital-heritage-leadership-briefing>.) Along with many other materials, Canadian archives include a large volume of orphan works (works for which the rights holder is unknown or unreachable) on a wide variety of topics and archivists and researchers would benefit greatly from more clarity on whether or not orphan works can be digitized to enable this kind of improved access. The newly created metadata that can be produced by generative AI will unlock access to the content of archival holdings that goes well beyond what is possible with the limited human resources currently available in archival institutions.

Recommendations:

- Provide clarity in the Copyright Act for the legal use of data for training generative AI tools
- Require AI researchers and developers to ensure training datasets have identifiable metadata that can be linked to generative AI output

Text and Data Mining

The Government of Canada invites views on whether any clarification is needed on how the copyright framework applies to text and data mining (TDM) activities, notably on how and when rights holders could or should be compensated for the use of copyright-protected content as inputs in the development of AI. Although all comments are welcomed, the Government is particularly interested in receiving feedback on the following questions:

- **What would more clarity around copyright and TDM in Canada mean for the AI industry and the creative industry? NA**
- **Are TDM activities being conducted in Canada? Why or why not? NA**
- **Are rights holders facing challenges in licensing their works for TDM activities? If so, what is the nature and extent of those challenges? NA**
- **What kind of copyright licenses for TDM activities are available, and do these licenses meet the needs of those conducting TDM activities? NA**
- **If the Government were to amend the Act to clarify the scope of permissible TDM activities, what should be its scope and safeguards?**

Libraries, archives, and museums (LAMs) are unlikely to be AI developers, so LAMs don't need an exception that permits them to use copies to train machines. LAMs are more likely to be asked to make copies of their holdings for AI developers upon request. If so, the fair dealing provision (s 29) of the Copyright Act (CA) may well serve, with some changes. The changes proposed below build upon provisions that are part of the balance between the rights of copyright owners and the interests of users already established within the CA.

Before describing the proposed changes, it is important to note that provisions such as fair dealing and the exceptions for LAMs are fundamental to the balance inherent in a well-functioning copyright system. Canada's approach to the challenges of AI must begin with established principles. The Supreme Court of Canada (SCC) has established that exceptions are not just loopholes, but users' rights (*CCH v LSUC* 2004 SCC 13 para 48), and we steadfastly defend their presence (particularly the fair dealing provision) as a fundamental principle. Since fair dealing is not limited to particular user groups, rights, formats, or categories of protected matter, everyone can benefit from it to access and use copyrighted material without authorization or payment, provided that the dealing is fair as determined by the SCC's two-step test.

As beneficiaries of fair dealing, LAMs already can make copies upon request for the purpose of research. Provided that TDM is appropriately defined to be clear that it is included within a broad and liberal interpretation of research, making copies for TDM falls within one of the allowable purposes of fair dealing. That uncertainty would be clarified if the fair dealing provision were amended by adding TDM or computational data analysis to the list of authorized purposes, OR by making the purposes illustrative rather than exhaustive, i.e., "fair dealing for purposes such as research, private study, ...do not infringe copyright." A further condition would require the LAM to inform the requester that the copies were provided for research only, that any further uses may require the permission of the

rights holder, and that it is the responsibility of the requester to obtain any necessary permissions. Admittedly, the scope of fair dealing may have to be clarified through litigation, since the limited case law cited in the consultation paper does not address situations where the copied images were used to train a machine.

Since users' rights are fundamental to a balanced copyright system, constraining them through contractual agreements undermines the system. Thus, the Copyright Act must be amended to provide that any contractual provision contrary to the exceptions in the Act shall be unenforceable.

The proposed amendments would provide legal clarity for both LAMs and AI developers by enabling LAMs to provide copies to AI developers to be used in the training of machine learning models.

Recommendations:

- Amend the fair dealing provision of the Copyright Act to provide that TDM lies within the scope of fair dealing.
- Amend the Copyright Act to provide that copyright exceptions cannot be overridden by contract terms.
- **What would be the expected impact of such an exception on your industry and activities? NA**
- **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?**

Having sufficient metadata that would identify and link the training datasets to the generative AI output created by the AI tool would be highly desirable.

Requiring AI developers to provide such metadata would provide transparency to the users of AI tools, and to the creators and rightsholders whose works are used in the datasets in a non-consumptive way. In order to ensure transparency and clarify rights issues, generative AI output should always be tagged as such.

Recommendation

- Require AI researchers and developers to ensure training datasets have identifiable metadata that can be linked to generative AI output.

- Generative AI output should always be tagged as such.

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

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

The possibility of a more general exception to permit TDM falls outside the scope of the archival community's direct interests. If, however, such an exception is needed, the provisions of Singapore's Copyright Act pertaining to computational data analysis (sections 243-244) are well thought out in terms of scope and appropriate safeguards. Its strengths are:

- Definition of "computational data analysis" (s. 243)
- Limited purpose (only Computational data analysis) (s. 244)(2)(a) & (b))
- Copy supplied/communicated to another only in very limited circumstances (s. 244)(2)(c) & 244(4))
- User must have lawful access to source materials (s. 244)(2)(d))
- Infringing source materials can be used subject to specific limited conditions (s. 244)(2)(e))

Authorship and Ownership of Works Generated by AI

The Government of Canada invites views on how the copyright framework should apply to AI-assisted and AI-generated content. Although all comments are welcomed, the Government is particularly interested in receiving feedback on the following 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?**

In Canada the basic principle is well established that copyright is automatic for original creations that include human skill and judgement, as specified in the Supreme Court decision (CCH Canadian Ltd. v. Law Society of Upper Canada, 2004 SCC 13, para. 25). The output that results from the generative

AI mechanical process cannot meet this requirement for skill and judgement and is therefore not protected by copyright. The humanly created algorithm does meet the requirement and is protected by copyright.

The current principle of not assigning copyright protection to generative AI output does not at all appear to be limiting the rapid development and adoption of AI technologies. The lack of certainty is, however, having a profound effect on creators and how they view their future prospects from both an economic and social standpoint.

- **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 believe that assigning full intellectual property rights to the output of generative AI processes is inappropriate. LAMs have a long history of advocating for clarity in the Copyright Act and we believe this issue must be addressed in the legislation, to provide as much clarity as possible.

Even with the current constraints and uncertainties, AI is profoundly disruptive in many ways, particularly to the creative communities. Assigning copyright protection to AI output would very negatively affect the work of creators and their contribution to society, resulting in a negative effect on incentive to create. Extending copyright protection to AI output calls into question the value we place on human creativity and expression.

AI processes can be programmed to create mass output that could quickly monopolize the creative space, thereby disrupting in profound ways human creative activity, the copyright balance, and the marketplace.

The rapid development and dissemination of AI has already created considerable disruptions to the creator community, and these will continue to be a major problem. Creators contest that the ingest of their works in the creation of the AI training models without attribution, permission, or financial compensation is a serious problem that will affect them in many

ways. But fair dealing and/or a TDM exemption would permit data mining for research purposes of the millions or billions of documents in the data sets used for training.

The prospect of directly compensating creators within the structure of the Copyright Act raises many thorny problems. Copyright law should not be used to address broad societal problems and challenges. However, copyright law is not the only way that we can reward creators. We recommend that the Government create a system outside the copyright regime to reward and acknowledge creators for the part their work plays in generative AI, such as a program in which AI developers are required to contribute to a fund that will be plowed back into the creator community to support a broad spectrum of Canadian creativity. (Other examples of this type of scheme are Canada's Public Lending Rights, Telefilm). The details of how such a program would work would have to be carefully considered, with input from the creator community, and the outcomes would have to include mandatory contributions by those developing the training datasets, and money paid out to the creator community. This would help redress the balance between human creators and the potential dominance of large corporate AI in the marketplace and the creation landscape. Such a program would enhance Government efforts to ensure support for Canadian creators and creative industries, while simultaneously fostering Canadian AI competitiveness, innovation, and support for maintaining overall access to Canadian creation, all of which are important public policy objectives.

Recommendations:

- Amend the Copyright Act to maintain and clarify the basic principle that copyright protects original creations that are the product of *human* skill and judgement and that the mechanical generative AI output is in the public domain, but the humanly created algorithm is protected by copyright.
- Create a system outside the copyright regime, that rewards and acknowledges creators for the part their work plays in generative AI, whereby generative AI developers are required to contribute to a fund that will be plowed back into the creator community to support a broad spectrum of Canadian creativity.

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

With further study and careful consideration, it is possible to consider very limited rights for AI outputs in particular circumstances, a variation of what is sometimes referred to as “thin copyright”, such as the limited rights sometimes accorded to databases. But these should be very limited in both scope and duration.

Recommendation:

- Consider very limited rights for AI outputs in particular circumstances, a variation of what is sometimes referred to as “thin copyright”.

Infringement and Liability regarding AI

The Government of Canada invites views on questions about copyright infringement and liability raised by AI, particularly since there is a lack of evidence currently available in this regard. Although all comments are welcomed, the Government is particularly interested in receiving feedback on the following 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)? NA**

What are the barriers to determining whether an AI system accessed or copied a specific copyright-protected content when generating an infringing output?

At present, there is no requirement for AI developers to provide metadata that would identify and link the training datasets to the generative AI output created by the AI tool. Requiring AI developers to provide such metadata would assist in determining whether protected material had been copied when generating an

infringing output, in addition to providing transparency to users of AI tools, and to the creators and rightsholders whose works are used in the datasets.

Recommendations

- Require AI researchers and developers to ensure training datasets have identifiable metadata that can be linked to generative AI output.
- **When commercialising AI applications, what measures are businesses taking to mitigate risks of liability for infringing AI-generated works? NA**
- **Should there be greater clarity on where liability lies when AI-generated works infringe existing copyright-protected works?**

It is clear that there should be greater clarity on where liability lies when AI-generated works infringe copyright-protected works. The current liabilities for copyright infringement should apply, but the issues will be clarified through litigation. Resolving the potential continuum of responsibility that arises with the actual situations in the litigation will be a more realistic approach, rather than rushing into a legislative solution that may have unintended consequences. The solutions must be consistent with the public policy issues discussed in other sections of this questionnaire

Recommendations

- Continue to apply current liability provisions and remedies for copyright infringement.
- Resolve liability and remedies issues that arise with ongoing litigation, to be consistent with sound public policy
- **Are there approaches in other jurisdictions that could inform a Canadian consideration of this issue? NA**

Comments and Suggestions

Are there any other considerations or elements you wish to share about copyright policy related to AI?

Diagnose and Understand the Problems

In this submission, we have pointed out some areas of deep concern, but have not necessarily proposed concrete or well-developed proposals of how they could or should be addressed. Expedient short-term solutions are not always the best way to proceed with problems arising with rapid change, as the hasty legislative “solutions” can create a whole set of new and unintended negative consequences. Diagnosing and understanding the problems is the important first step in finding viable solutions and this requires time and careful consideration.

Further Consultation Must extend beyond Copyright Stakeholders. We believe that it is crucial that public policy and legislation concerning Artificial Intelligence should be undertaken only after wide public consultation and discussion that includes broad public policy concerns such as protection of privacy and personal information, and other human rights. It is clear AI is introducing significant disruption to current realities, and it is important to remember that this disruption is taking place not only in the commercial marketplace, but also in research and in many other spheres of public life. Across the globe, we are struggling to understand the impacts of this rapidly evolving technology. Despite its many benefits, there are many public interest issues that are of concern to LAMs, concerns that go well beyond copyright—for example machine learning bias, misinformation, privacy issues, data breaches, and protection of freedom of expression. The Archival Community believes that all these issues require careful examination and public consultation that extends beyond this consultation and its copyright concerns. We believe it’s just too early to know how to deal with some of these complex and rapidly emerging problems.

Any changes to policy and legislation, be it copyright or other initiatives, must safeguard the public interest and privacy considerations. Effective public consultation with public interest groups and the broad General Public on the complex issues surrounding AI must take place. It is not only copyright

stakeholders and those with a vested interest in the marketplace who should be part of the discussion. Copyright concerns for AI must not be addressed in isolation and without reference to broad public concerns, including the professional, economic, and social disruptions that come with AI. These are extremely important issues for archivists because of the nature of our holdings and our public service mandate. Broad public consultation and discussions are required to develop sound public policy and legislation that will avoid decision-making that is solely market-oriented. The public good will be served by such consultations that are in line with Open Government policies and processes already in place in the Canadian government.

AI is developing at a galloping pace, but we must avoid knee-jerk reactions that do not take into consideration the broad range of public policy issues that are intricately connected to this emerging technology.

Only Litigation Will Solve Some Problems We believe that some of the thorny issues around AI will only be clarified through litigation. This is inevitable in any environment of rapid change. Allowing the litigation to follow its course may be helpful in seeing clearly where the problems are.

Recommendation:

- Extend the consultation and discussion of these issues to the General Public to ensure that public policy issues are considered in decision-making concerning AI policy and legislation.