Canadian Federation of Library Associations
Fédération canadienne des associations de bibliothèques

Consultation on Copyright in the Age of Generative Artificial Intelligence: CFLA Response

Technical Evidence

Libraries, archives and museums (LAMs) support AI research in the development of training datasets for use in AI models, particularly those used to train language models. Libraries provide access to large corpora of text and facilitate the licensing of content for AI purposes. Canadian university libraries informally report that researchers are stymied by scholarly publishers' poor tools and high licensing costs for AI research. These tools are expensive, proprietary, and lack the functionality researchers need. Licensing costs for TDM activities are now a revenue stream for large multinational publishers, requiring libraries to pay multiple times for use, albeit different uses, of the same content. Such actions exemplify the drive to commodify all uses and thereby shrink the commons, threatening the public good and upsetting the Copyright Act's balance between users and rightsholders.

Some publishers block the non-consumptive use of published works for AI training, while at the same time collecting data and usage patterns of their paying customers to develop AI systems for further commercial purposes (Yoose & Shockey, 2023), threatening privacy and equity standards. Researchers often need access to a wide variety of data sources in order to protect against bias so high costs and extra licenses needed for TDM access can inhibit research.

Libraries are centres of copyright expertise within many organizations, and are called upon by researchers to provide assistance in understanding the copyright implications of an AI research project. Researchers and librarians want to ensure the responsible development of AI and this includes ensuring that copyright is considered and respected. Most potential training datasets are not neatly packaged up, analyzed for copyright issues, and made available under a legally vetted licence. Instead, most training datasets are either vast in size, custom built for training a specific

model, or for transfer learning. (Transfer learning is common in LAMs and occurs when researchers use an already trained model and then introduce a small new dataset in order to refine the model so that it better accomplishes a specific task.) Consequently, the use of most training datasets requires a fair dealing assessment in order to mitigate the risk of infringement. Librarians provide copyright guidance to researchers on their proposed use of training datasets, and their use of generative AI systems to create new works. This guidance is needed as the current formulations of sections 29 and 30.71 of the Copyright Act lack clarity for a researcher to know if the training of AI models with their proposed dataset is copyright infringing. Clarity through a specific exception would assist researchers in their AI projects as well as libraries in providing copyright guidance.

In libraries and educational institutions, human input is significant in the development of AI models and datasets. Many developers practice human centred explainable AI, centring the human in AI development, letting us understand and contest generative AI outputs and the decisions underlying those outputs (Ehsan et al., 2023). Thus, to mitigate bias in generative AI models, we need diverse and inclusive datasets. Market solutions providing datasets that are curated and licensed by rights owners is insufficient and public domain materials and openly licensed materials lack sufficient diversity for bias-reduced training of AI models. Thus AI models must be trained on all kinds of works, including unlicensed copyrighted works.

Technological neutrality helps navigate the copyright implications of using datasets containing copyrighted works to train generative AI models. Both the Summary to the Copyright Modernization Act and Supreme Court jurisprudence reminds us of the importance of technological neutrality in preserving the balance between authors and users in the digital environment (Entertainment Software Association v. Society of Composers, Authors and Music Publishers of Canada, 2012, paras 7-8). Technological neutrality implies that Canada's core copyright understandings must be consistently applied "in a manner that appropriately balances the rights and interests at stake - maintaining in the face of technical change, the steady pursuit of copyright's policy goals" (Craig, 2017, p. 612).

Generative AI is an evolving technology which enables the analysis and production of information at a speed and scale impossible for human beings. Such technology disrupts our current copyright framework and raises questions about how, or if, this

technology implicates the exclusive rights of copyright owners. However, using the lens of technological neutrality allows for copyright to adapt to new disruptive technologies and lets us “maintain normative vigilance as conditions change” (Craig, 2017, p. 617) rather than constantly extending copyright owners’ exclusive rights when the activities of new technologies do not actually engage the copyright owner’s legitimate interests. A work that is copied to be reduced to a collection of discrete elements, or underlying facts and ideas, for training an AI model is not copied for human enjoyment and is not engaging with the author’s interests nor with an incentive to create. A technologically neutral functional equivalence approach tells us that copies made for training AI models do not implicate exclusive rights. To argue otherwise risks entertaining the concept that the acts of reading and memorization of works engages exclusive rights in an infringing way.

There is concern in the research community about training data and mitigating the risk of copyright infringement, but also about ensuring transparency and non-bias in training data. Many of these same researchers are concerned about the impact of the generated products on rightholders and are working on solutions to attribute, or link, training data to the generated works to provide greater transparency to the user. To do this effectively will require that training datasets properly identify the source of each discrete element of content in the dataset.

Canadian LAMs use generative AI tools in multiple ways. For example, university libraries and archives are using computer vision AI and generative AI tools to create extensive metadata for existing analogue image collections. LAMs also use generative AI models to create basic metadata for each document in large scale digitization projects.

Generative AI holds out great promise to enable libraries and archives to provide new access points and greater descriptive metadata to their collections than is currently possible. Generative AI transcription tools, such as Whisper, when trained with specific datasets incorporating content from the collections of libraries or archives, extracts information from audio files (e.g. oral histories and interviews) about the subject matter and the people involved. For example, these tools can extract the titles of all the poems

recited and the types of questions asked by the audience in a poetry reading recording; this type of description is too labourious and time consuming without the aid of AI.

Film and media archives use generative AI to move beyond simple descriptions and allow researchers to engage with film in ways we never did in the past, and assist with a wide range of accessibility needs (Mason, 2023). These endeavours are too time consuming for humans to carry out, but generative AI makes it feasible for libraries and archives to provide rich metadata and vastly increased discoverability and access to collections. When utilizing generative AI tools for enriching descriptions and access, the original works are usually in analogue format. Therefore these works need to be converted - copied - into a digital format so that they can separately be ingested into the AI system for individual analysis. These copies are not necessarily for TDM purposes or for dataset training purposes, nor are they being made under the preservation and obsolete format provisions of the Copyright Act. For libraries and archives to use generative AI tools to enhance discoverability and access, they must be confident that the copies they make to utilize the promise of generative AI are not considered compensable or infringing.

Recommendations:

1. Provide clarity around training dataset content by encouraging training datasets to have sufficient metadata such that each content element is identifiable.
2. Ensure that any Copyright Act exception for the creation of non-consumptive copies for the purpose of informational analysis is broad enough to allow LAMs and other users to make non-consumptive copies of works. This would include the ability to circumvent a TPM to make such copies, or for purposes of utilizing technological tools such as generative AI to create metadata and enable superior discovery of those works.

Text and Data Mining

Text and data mining (TDM) involves the automated identification of patterns within vast datasets, playing a crucial role in the advancement of artificial intelligence (AI). TDM entails creating non-consumptive duplicates of materials, some of which may be subject to copyright (Non-consumptive copies are copies that are utilized for purposes other than the works' original objective (i.e. reading, studying, performing, etc.), but are used

for technological purposes such as web caching, or for data processing purposes, like TDM). The legal status of TDM currently lacks clarity, and the absence of a specific

TDM exception in the Canadian Copyright Act hinders researchers' efforts and impedes progress by requiring extensive copyright analysis to ensure compliance. The comments below build upon our community's previous submissions and statements, which offer additional examples illustrating the crucial role that libraries play in this domain(CFLA, 2023; CFLA, 2018; Portage, 2018). Many of the questions that TDM analysis pose are central to larger issues that libraries are struggling with as our collective works are moving from traditional formats such as print, in which we could rely on copyright laws and exceptions, to digital access where fundamental user rights are quickly eroded under licensing terms and weakened by technological protection measures (TPM). The suggestions and remarks provided below are embedded into the wider framework of safeguarding the overarching goals of promoting fair access to knowledge and information for the 'public good' (Liber, 2020; IFLA, 2020). These goals are directly impacted by any alterations Canada may make to its copyright legislation to responsibly and equitably address technological advancements such as Generative AI, that uses TDM.

It is crucial to recognize that generative AI and TDM analysis are distinct tools. TDM is an analytical tool which involves the automated identification of patterns from extensive datasets. Certain applications of this method of analysis involve a large corpus of textual data. It is important to articulate that the library community support for TDM is based upon applications of this technology that are not concerned with or attempting to encroach on the vested copyrights of the original expression of a work, but to facilitate analysis that unearth patterns, information, and correlations, from the facts and ideas behind these works. Librarians and archivists believe this non-expressive, or non-consumptive, use of a work should be protected in copyright legislation through an exception. Any limitations or regulations applied to TDM will have a significant impact on the future shape and value of generative AI among other forms of analysis of digital works.

As has been stated in previous submissions (CFLA, 2021), the library community is familiar with the limitations and chilling effects that current copyright legislation imposes. Libraries are finding efficiencies and technologies to keep up with the proliferation of all

formats of works (National Lottery Heritage Fund, 2023). However, restrictive licensing terms of use, digital locks and technological protection measures erode well-established user rights and inhibit access. In an example of how such restrictions can affect scholarly work, a Canadian-led group of researchers was forced to retract a paper that had been accepted for publication on vaccine hesitancy and COVID-19 because, while the law would allow it, the database contract overrode the statutory rights of the researchers; they had not secured a licence to mine a database of news articles used in the study (RetractionWatch, 2021; CFLA, 2023). Libraries acknowledge the need for mechanisms that would allow for and incentivize a market for TDM data, but these incentives must not come at the expense of basic user rights to the original publication or access to the facts and data of the expression. The non-consumptive nature of these analytical uses of works is an important concept to build into any technologically durable copyright policy.

As outlined in this Consultation's Paper, there are two general directions that address TDM within copyright legislation in other jurisdictions. The library community supports the introduction of a specific TDM exception. This approach provides a practical basis for users and a solid framework for libraries to support research and creativity; however, we caution against overly restrictive language potentially leading to unexpected

obstacles as technology and expression evolve. A number of Canada's key trading partners already have a specific exception for TDM, including Japan,

Singapore, United Kingdom, and the EU. The library community supports an exception that applies to both commercial and non-commercial research, that includes both the reproduction right and communication right. Japan's 2018 TDM exception, based on Article 30-4 of its Copyright Act, specifies that non-consumptive copies do not infringe upon the rights of the copyright owner. This Japanese exception permits TDM for both commercial and non-commercial purposes and prohibits rights holders from making TDM reservations(Ueno, 2021). Additionally, it nullifies contractual clauses attempting to restrict TDM.

Libraries should be able to override contract restrictions that thwart statutory rights and copyright exceptions so that vendors cannot make TDM reservations and/or fair dealing reservations. Singapore's 2021 TDM exception also allows for both commercial and non-commercial TDM, explicitly forbidding contractual overrides. Moreover, similar to

Singapore's 2021 Computational Data Analysis amendment, this exception should equally apply not only to Canadian law-governed contracts, but also contracts governed by foreign law "where the choice of foreign law is wholly or mainly to evade any copyright exception"(Kang, 2021).

To safeguard the integrity of the balance of user rights a TDM exception needs to be supplemented with illustrative language within the fair dealing framework, by adding the words "such as" to the purposes given in S.29 of our Act that will allow users to confidently apply basic user rights across creative expression (CFLA, 2023). For example, the use of illustrative language in the US has established a solid legal basis within their fair use framework for non-consumptive research on copyrighted materials for uses such as TDM. Legislation that anticipates fair and diverse access to information requires an approach that does not over-inflate the expressive capabilities of machine generated output or undervalue the importance of access to the widest possible scope of information that will enable unbiased applications of this form of analysis.
Contrary to licensing as a viable solution for TDM, libraries argue, as articulated by the International Federation of Library Associations (IFLA), that the right to access content should inherently encompass the right to engage in text and data mining. As stated by IFLA:

> [T]he right to read ... content should encompass the right to mine. Further, the sheer volume and diversity of information that can be utilized for text and data mining, which extends far beyond already licensed research databases, and which are not viewed in silos, makes a licence-driven solution close to impossible (IFLA, 2013).

Since research is often conducted by international teams, CFLA recommends that an international instrument for TDM be developed at WIPO to ensure that cross border research is not hampered by a patchwork of national legislative barriers. The vast and diverse range of information available for text and data mining, extending beyond licensed research databases and not compartmentalized, makes a license-driven solution nearly impractical.

Recommendations

1. Create a specific exception for TDM. The library community supports the creation of a specific exception that would "facilitate the use of a work or other subject-matter for the purpose of informational analysis".

2. Further facilitating TDM: prohibit contract override and allow circumvention of TPMs for any non-infringing purpose. CFLA recommends introducing an exception that prevents contracts from overriding copyright exceptions for non-infringing purposes. This provision should apply to all future and pre-existing contracts.
3. Make fair dealing purposes illustrative. CFLA supports recommendations in the 2019 Copyright Review related to the enumerated list of purposes under Section 29 of the Copyright Act.
4. Support the creation of a specific international exception for TDM.

Authorship and Ownership of Works Generated by AI

The current Copyright Act has achieved a certain balance that would be disrupted by including AI outputs (CFLA, 2023). The current Copyright Act has achieved a certain balance that would be disrupted by including AI outputs. The Copyright Act safeguards works crafted by human authors, including the underlying computer programs of AI. The development and adoption of AI technologies is not inhibited by the current lack of copyright protection of AI-generated works. However, the lack of a policy framework for generative AI is having an impact on creators, and could be addressed in a number of ways outside of copyright.

In Canada, copyright serves to protect the expression of human creativity, encompassing both skill and judgment. Outputs from mechanical and routine processes do not meet the originality standard set by the unanimous CCH decision of the Supreme Court of Canada (CCH Canadian Ltd. v Law Society of Upper Canada, 2004). Without expressive agency and intellectual effort, the outputs of AI processes should not be accorded similar copyright protection as works by human creators. Carys Craig underscores that "authorship involves expressive agency, a quality inherently lacking in AI (Craig, 2021a)." Granting machines the status of rights holders is contrary to the current provisions in the Copyright Act.

The outputs of AI processes without significant human intervention represent mechanical exercises devoid of skill and judgment, contrasting with the exercise of skill and judgment in developing an algorithm. Consequently, a computer program is protected by the Copyright Act. Unlike human authors, AI processes do not rely on

copyright incentives to produce new works (Gervais, 2020). Expanding protection of intellectual property rights to outputs generated by AI machines could upset the balance of IP protection and discourage other stakeholders.

AI processes possess the ability to generate works more rapidly and systematically than human authors. The substantial output facilitated by AI has the capacity to displace human creativity and introduce economic disruptions, disadvantageous to human

authors while favoring the swift and serendipitous outputs of machines. One of the primary purposes of copyright is to strike a balance between the rights of authors and the broader public interest, particularly in education, research, and access to information (WIPO, 1996). The extensive output enabled by AI has the potential to disrupt the economy by placing human authors at a disadvantage. If subjected to a comprehensive spectrum of copyright protections, this volume-driven "autoship" could marginalize human authors' outputs and undermine society's right to access facts and information that would otherwise remain in the public domain. Giving the full duration of copyright protection to AI-generated works could result in copyright overreach on a massive scale, allowing some AI companies the potential ability to crowd out human creators in such areas as music (Obeebo Inc., 2019).

On the matter of authorship, CFLA currently advocates that outputs of AI processes remain unenclosed and open to the public. As cautioned by Craig and others, extending full copyright protection to AI outputs poses a threat to the equilibrium of copyright and challenges the value Canada places on human expression (Craig, 2021a; Copyright Review Board, United States Copyright Office, 2022.)
The Copyright Act should remain as is with regards to human authorship. Granting copyright protection to artificial intelligence (AI) outputs could disrupt the intricate and nuanced equilibrium established in the Copyright Act. The Act currently safeguards works crafted by human authors, encompassing the computer programs that form the foundation of AI.

In circumstances where sufficient human expressive agency is added to an AI-generated work (e.g. the output of a generative AI process has been substantially re-edited in Photoshop), as the U.S Copyright Office notes works could be afforded copyright protection under certain circumstances (United States Copyright Office, March 2023). The granting of any such protection would be judged on a case-by-case basis,

and copyright ownership claims could be based on documentation of the exercise of human skill and judgment dedicated to the revision of the AI generated work. Possible ways to make clear the origins of AI-generated works include the addition of metadata that identifies the work as AI-generated. For example, the private company

Stability AI is currently working on a tool that will tag image content generated with their tool with metadata that discloses the AI origin of the work, which could be protective both to distinguish AI generated work from human expressive content and also protect against “deep fakes (Stability AI, 2023).” If AI generated works are marked in some way it will be easier to trace the public domain copyright status of AI outputs, to distinguish them from works that have copyright protection.

The Canadian Intellectual Property Office (CIPO) should refrain from granting copyright registration to AI created works, and refrain from acknowledging AI machines as co-authors or single authors of works. CIPO should be guided by the work done by the Copyright Office in the United States, which produced “Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence in March 2023 (United States Copyright Office, 2023).” This document makes it clear that human authorship is necessary for copyright registration and protection, unless significant human input is made to the resulting output.

Recommendations

1. Artificial intelligence authored works should not be protected by copyright.
2. The Canadian Intellectual Property Office (CIPO) should refrain from granting copyright registration to AI created works, and refrain from acknowledging AI machines as co-authors or single authors of works without the applicant showing significant human input has been made to the outputs.
3. While CFLA recognizes that generative AI can be disruptive to creators, the issue of possible compensation for creators of material used to train AI machines should be separate from the Copyright Act.

Infringement and Liability regarding AI

CFLA’s position is that non-consumptive copies of works used to train AI machines that are obtained via TDM, are allowable uses under the current copyright exception fair

dealing. CFLA recommends that a TDM exception be added to the Copyright Act. Please refer to our Text and Data Mining recommendation on contract override provisions and adding illustrative purposes to fair dealing. AI brings with it a host of opportunities for misuse and the exploitation of the power of AI to pursue illegal endeavours such as privacy intrusions, large scale copyright infringement, illegal collection of data and other actions. These issues are significant from a public policy perspective, and they must be addressed separately with the copyright implications of AI.

The Copyright Act in its current form does provide legal remedy for copyright infringement in works created through generative AI. If an AI-generated output is determined to be substantially similar to an already existing human created work, it can be subject to a copyright infringement claim, which would be decided through the courts. This does not need additional clarification in the Copyright Act. The question of who would be considered liable (e.g. the entity providing access to the generative AI product or service, the programmer or the end user) exists on a continuum and would need to be evaluated on a case-by-case basis through the courts. We reiterate generative AI outputs should not have copyright protection and should remain in the public domain. Courts need to also protect against copyright misuse (Twigg, 2012), so rights holders seeking protection in areas such as style, ideas, facts and data do not overreach the statutory limits of copyright, and encroach on the public domain. However, since derivative works created by AI could be influenced by a number of factors, including a dearth of training data, infringement by users of AI services in many circumstances may be unintentional. As well, since some generative AI services such as ChatGPT also disclaim responsibility for similarity to content that is produced by their tools, they do not guarantee that similar material might not be created for multiple users of their service (OpenAI, 2023). Incidental copyright infringement might also be recognized as a possibility of defense when it comes to accidental infringement in AI-generated works, as at the moment for many AI generated artworks for example, outputs are somewhat random, and keywords used can only guide outputs. In Section 30.7 of the Copyright Act, the “incidental inclusion” provision is worded as follows: “It is not an infringement of copyright to incidentally and not deliberately: (a) include a work or other subject-matter in another work or other subject-matter; or (b) do any act in relation to a work or other subject-matter that is incidentally and not deliberately included in another work or other subject-matter (RSC,1985, c. C-42).”

Libraries would like to see the ongoing development and use of generative AI centring transparency in how AI models are trained, algorithms are used, and the design and intentions behind AI tools. This kind of transparency is essential to protect and inform users about how generative AI tools make decisions, especially when it comes to certain applications such as in healthcare. This transparency goes far beyond creator and copyright issues in terms of impact on Canadians.

A lack of transparency when it comes to training data can be an obstacle for the discovery of if a non-consumptive copy of a specific copyright protected work was used in the AI-training process. Using metadata tags to track training material could, in part, help remedy this situation. AI developers should keep records of where training data came from in machine training, and be required to disclose training data summaries in response to claims of infringement. Infringement claims should be based on the similarity of AI-created outputs to training materials that have been ingested, not based merely on non-consumptive copying of content, and infringement should be decided in a court of law. However, transparency requirements need to remain flexible, not be retroactive, and allow sufficient time for AI developers to plan for and implement.
Some private generative AI companies that used creative copyright-protected works to train their machines, such as Stability AI, have already taken steps to create tools that allow for creators to opt-out of the inclusion of their work in the companies' models going forward (Heikkilä, 2022). The ability to opt-out of training data however should remain a private ordering, and not be legislated. Legislating TDM, so as to allow opt-outs could have a number of significant unintended consequences. By limiting the potential sources of data on which AI tools can be trained, it could contribute further to existing issues of bias and inequality in AI-generated outputs as well as having serious long term effects on the future reliability of AI machines in certain applications such as health care, autonomous vehicles, etc (Craig, 2021b, p.3; Creative Commons, 2021, p.6).
Copyright infringement liability should be determined on a case-by-case basis in the courts. Liability for copyright infringement when it comes to outputs could either lie with the developer, the AI company, or the user, and could lie on a continuum. Additionally, liability when it comes to generative AI goes far beyond copyright when it comes to "high risk" applications for consumer use such as medical uses or self-driving cars, and the evaluation of whether user error or a defect was present in the design of the AI when it was released (Long, 2023).

The threat of liability will have an impact on cultural heritage institutions that are mandated to preserve, disseminate, and provide access to knowledge, culture, and history. These public good institutions need clear protection from liability so that they can continue their mission.

As there is so little case law in the area of liability many jurisdictions may be using a "wait and see approach" before moving forward with legislation, and Canada may be wise to follow suit (Congressional Research Service, 2023).

The AI Act in the European Union offers some guidance. It stipulates that any image, audio, or video content displaying a noticeable similarity to authentic or truthful content i.e ('deep fake') must be revealed as having been generated through automated means unless it is for some allowable purposes (European Commission, 2021). While still to be determined, it may be useful to have identification mechanisms such as metadata for some AI generated creative content to identify outputs created via generative AI in order to distinguish them from copyright protected works, as well as having the ability to identify "deep fake" works (Barney & Wigmore, 2023).

As mentioned in the Text and Data Mining section of this response, a number of Canada's key trading partners already have a specific exception for TDM, including Japan, Singapore, United Kingdom, and the EU. The library community supports an exception that applies to both commercial and non-commercial research, that includes both the reproduction right and communication right such as Japan's 2018 TDM exception, based on Article 30-4 of its Copyright Act, specifies that non-consumptive copies do not infringe upon the rights of the copyright owner (Ueno, 2021). This Japanese exception permits TDM for both commercial and non-commercial purposes and prohibits rights holders from making TDM reservations (Ueno, 2021). Additionally, it nullifies contractual clauses attempting to restrict TDM. Singapore's 2021 TDM exception also allows for both commercial and non-commercial TDM, explicitly forbidding contractual overrides (Kang & Oh, 2021).

Recommendations:

1. Please refer to our Text and Data Mining section for recommendations in this area.

2. A mechanism for judging copyright infringement for generative AI outputs already exists in Canadian copyright laws, and whether a use is infringing should be determined on a case-by-case basis in the courts.
3. Liability when it comes to AI outputs that are potentially infringing could either reside with the developer, the AI company, or the user, and could lie on a continuum.
4. There may need to be a consideration of incidental inclusion when it comes to generative AI outputs.
5. As there is so little case law in the area of liability many jurisdictions may be using a “wait and see approach” before moving forward with legislation, and Canada may be wise to follow suit.

Summary

Copyright law should not be utilized as a tool to tackle the broader societal challenges that may result from the effects of generative AI on society. Nor should AI innovation be constrained in Canada by laws that are inflexible and have fewer exceptions than other competing jurisdictions, such as the US, which has an expansive fair use doctrine for AI developers and researchers to rely on.

AI possesses the capacity to revolutionize numerous occupations beyond individual creators, and such disruptive innovations have been seen throughout human history such as the printing press, automation in industry, and the digital disruption of the internet, to name a few examples. Addressing the resultant innovative disruption by supporting training for new opportunities in jobs related to AI development or by supporting worker retraining through organizations like community colleges, universities

and public libraries, should be approached at an economic and society wide level (Library Copyright Alliance, 2023). As well, the Canadian government should invest in more grants and support for Canadian creative industries and for creators in the long term.

As it currently stands there is a huge swath of information that is unavailable to Canadian higher education researchers and smaller independent AI researchers because of technological protection measures and prohibitive licensing fees to access some data sets. This includes licensed library resources that in many cases require additional text and data mining agreements to be able to be used by institutional

researchers for TDM purposes. Researchers may need access to many sets of data in order to complete a project, and there is a real risk that these research projects might not be realized. There is a societal risk of a regime of monopolistic access to data, where large AI or data companies are the only ones that can afford to gather, purchase or assume the risk of accessing data sets (Internet Archive, 2023). Democratic access is reduced under licensing regimes. It is in the public interest for Canadian AI researchers to have robust exceptions when it comes to TDM.

References

Barney, N., & Wigmore, I. (2023, March 21). What is Deepfake AI? A definition from TechTarget. WhatIs. https://www.techtarget.com/whatis/definition/deepfake

CARL-ABRC Fair Dealing Comparison Chart (2018). Retrieved from https://www.carl-abrc.ca/wp-content/uploads/2018/07/Fair-dealing-comparison-chart.pdf

CCH Canadian Ltd. v Law Society of Upper Canada, 2004 SCC 13, [2004] 1 SCR 339.

CFLA-FCAB Brief to the Government of Canada Consultation on a Modern Framework for Artificial Intelligence and the Internet of Things (2021). Retrieved from https://cfla-fcab.ca/wp-content/uploads/2022/01/CFLA-CARL-Brief-Artificial-Intelligence-and-the-Internet-of-Things.pdf

CFLA-FCAB Statement: Copyright and Text and Data Mining (TDM) Research (2023). Retrieved from https://cfla-fcab.ca/wp-content/uploads/2023/07/CFLA_FCAB_Statement-on-Text-and-Data-Mining-Research-1.docx.pdf

CFLA Copyright Committee. (2023). CFLA Statement: AI and Copyright and its application in Cultural Heritage Institutions. https://cfla-fcab.ca/wp-content/uploads/2023/07/CFLA-FCAB_Statement_on_AI__Authorship-1.docx.pdf

Congressional Research Service. (2023, September 29). Generative Artificial Intelligence and Copyright Law. (Legal Sidebar LSB10922) https://crsreports.congress.gov/product/pdf/LSB/LSB10922

Copyright Act, Revised Statutes of Canada (1985, c. C-42). https://laws-lois.justice.gc.ca/eng/acts/C-42/

Craig, C. J. (2017). Technological neutrality: Recalibrating copyright in the Information Age. Osgoode Legal Studies Research Paper Series, 186. http://digitalcommons.osgoode.yorku.ca/olsrps/186

Craig, C. J. (2021a). AI and Copyright., in Florian Martin-Bariteau & Teresa Scassa, eds., Artificial Intelligence and the Law in Canada. Toronto: LexisNexis Canada.

Craig, C. J. (2021b). Joint Submission of IP Scholars, Re. Consultation on a Modern Copyright Framework for Artificial Intelligence and the Internet of Things" . Commissioned Reports, Studies and Public Policy Documents. Paper 226. https://digitalcommons.osgoode.yorku.ca/reports/226

Creative Commons (2021). Submission by Creative Commons. Government of Canada Consultation on a Modern Copyright Framework for Artificial Intelligence and the Internet of Things. https://wiki.creativecommons.org/images/f/f6/Creative_Commons_submission_Canada_consultation_AI_and_the_internet_of_things.pdf

Copyright Review Board, United States Copyright Office. (2022, February 14). Second Request for Reconsideration for Refusal to Register A Recent Entrance to Paradise. (Correspondence ID 1-3ZPC6C3; SR # 1-7100387071).”

Ehsan, U., Wintersberger, P., Watkins, E.A., Manger, C., Ramos, G., Weisz, J.D., Daum, H., Reiner, A., Riedl, M.O. (2023). Human-centered explainable AI (HCXAI): Coming of age. Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems. https://doi-org/10.1145/3544549.3573832

Entertainment Software Association v. Society of Composers, Authors and Music Publishers of Canada. 2012 SCC 34. https://scc-csc.lexum.com/scc-csc/scc-csc/en/item/9994/index.do

European Commission (2021, April 4). Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial

Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (COM/2021/206 final). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206

Gervais, D.(2020). The Machine as Author. Iowa Law Review. 105 (2062), 2053–2106.

Heikkilä, M. (2022, December 16). Artists can now opt out of the next version of Stable Diffusion. MIT Technology Review. https://www.technologyreview.com/2022/12/16/1065247/artists-can-now-opt-out-of-the-next-version-of-stable-diffusion/

IFLA. IFLA Statement on Libraries and Artificial Intelligence (2020). Retrieved from https://www.ifla.org/publications/ifla-statement-on-libraries-and-artificial-intelligence/

IFLA. IFLA Statement on Text and Data Mining. Retrieved from https://www.ifla.org/publications/ifla-statement-on-text-and-data-mining-2013/

Internet Archive. (2023, Nov. 1). Internet Archive's Public Comments in Response to the Copyright Office Study on Artificial Intelligence. Comment from Internet Archive. U.S. Copyright Office. https://www.regulations.gov/comment/COLC-2023-0006-8836

Library Copyright Alliance. (2023, July 10). Library Copyright Alliance Principles for Copyright and Artificial Intelligence. https://www.librarycopyrightalliance.org/wp-content/uploads/2023/06/AI-principles.pdf

Ligue des Bibliothèques Européennes de Recherche Association of European Research Libraries (LIBER). (2020). Text and Data Mining. The need for change in Europe. Retrieved from https://libereurope.eu/wp-content/uploads/2020/11/TDM-Copyright-Exception.pdf

Long, R.E. (2023, March 17). Artificial Intelligence Liability: The Rules Are Changing. The Center for Internet and Society. https://cyberlaw.stanford.edu/blog/2023/03/artificial-intelligence-liability-rules-are-changing-1

Kang, A. & Oh, P. (2021, September 20). Coming Up in Singapore: New Copyright Exception for Text and Data Mining. https://www.lexology.com/library/detail.aspx?g=1ce9c997-22a1-4953-bd0b-68a95d31bc89

Mason, I. (2023, November 15-17). Organizing within LAMs to address AI [Conference Presentation]. AI4LAMs Conference, Vancouver, BC, Canada.

National Lottery Heritage Fund. (n.d.). Artificial Intelligence: a digital heritage leadership briefing. The National Lottery Heritage Fund provides many examples of the use of AI in libraries, archives, and museums. Retrieved from]https://www.heritagefund.org.uk/about/insight/research/artificial-intelligence-digital-heritage-leadership-briefing

Obeebo Inc. (2019). Comments on Intellectual Property Protection for Artificial Intelligence Innovation., submitted to USPTO Request for Comments on Intellectual Property Protection for Artificial Intelligence Innovation. https://www.uspto.gov/sites/default/files/documents/Obeebo-Inc_RFC-84-FR-58141.pdf

OpenAI (2023, November 14). Terms of use. https://openai.com/policies/terms-of-use

Portage. (2018). Brief to INDU Committee on Text and Data Mining. Retrieved from https://www.ourcommons.ca/Content/Committee/421/INDU/Brief/BR10245456/br-external/PortageNetwork-e.pdf

RetractionWatch. (2021, July 30). ‘A very unfortunate event’: Paper on COVID-19 vaccine hesitancy retracted. Retrieved from https://retractionwatch.com/2021/07/30/a-very-unfortunate-event-paper-on-covid-19-vaccine-hesitancy-retracted/

Stability AI. (2023). Comment from Stability AI. U.S. Copyright Office. https://www.regulations.gov/comment/COLC-2023-0006-8664

Twigg, M. (2012). Copyright Misuse: Protecting Copyright in Canada from Overreach and Abuse. Dalhousie Journal of Legal Studies, 21(1). https://digitalcommons.schulichlaw.dal.ca/djls/vol21/iss1/6/

Ueno, T. (2021)The Flexible Copyright Exception for 'Non-Enjoyment' Purposes – Recent Amendment in Japan and Its Implication. GRUR International, 70 (2), 145–152. https://doi.org/10.1093/grurint/ikaa184

U.S Copyright Office. (2023, March 16). Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence. Federal Register, 16191-16192. https://www.federalregister.gov/documents/2023/03/16/2023-05321/copyright-registration-guidance-works-containing-material-generated-by-artificial-intelligence

World Intellectual Property Organization. (1996). WIPO Copyright Treaty. Geneva.

Yoose, B., & Shockey, N. (2023). Navigating Risk in Vendor Data Privacy Practices: An Analysis of Elsevier's ScienceDirect (Version v1). SPARC. https://doi.org/10.5281/zenodo.10078610