

Government of Canada Consultation on a Modern Copyright Framework for Artificial Intelligence and the Internet of Things

Submission by Creative Commons

Prepared by Brigitte Vézina, Director of Policy, Open Culture and GLAM, Creative Commons
September 17, 2021

Supported by [Creative Commons Canada](#)

Submitted by email: copyright-consultation-droitdauteur@canada.ca

Introduction

Creative Commons (CC) welcomes the opportunity to provide feedback to the Government of Canada's [Consultation on a Modern Copyright Framework for Artificial Intelligence and the Internet of Things](#). The present document represents CC's provisional position on the issues raised in the consultation. Our three key points can be summarized as follows:

- Copyright law should allow unrestricted access and use of protected works to train artificial intelligence (AI) at the point of input.
- AI-generated output should not be protected by copyright.
- Technological protection measures should not be used to control, limit, prevent or otherwise affect legitimate activities and uses allowed under a copyright exception or limitation regime as well as under CC licenses' terms.

Artificial intelligence

Developments in artificial intelligence (AI) present a host of exciting opportunities in the digital world. These range from the development of models or algorithms perfected through data processing, to mining, analysing and enriching datasets with new metadata. This is especially true for GLAMs (galleries, libraries, archives and museums) and the cultural heritage sector in general. While these opportunities are likely to propel GLAMs forward through their digital transformation, they also raise questions in the area of copyright, especially when it comes to using GLAMs' digital collections to train AI and the treatment of AI-generated outputs under copyright law.

Our comments are provided subject to the following caveats regarding the uncertainty around AI and its rapid evolution:

- Clarity on basic definitions in the AI space is a prerequisite to competent regulation in the copyright arena. AI needs to be properly understood before any copyright implications can be addressed.
- Any policy or legal intervention in the field of copyright should be based on strong and reliable evidence and conceptual certainty, especially given the fast-paced evolution of AI technology.

Use of copyrighted content as input for AI training

CC fully supports broad and unfettered access and use of data to train AI. In the GLAM sector, cultural heritage institutions must be able to use the massive amounts of data in their digital collections for AI-training purposes (including machine-learning) in order to [fulfil their public interest missions](#). Legally, there remains significant uncertainty as to whether copyright limitations and exceptions allow the use of copyright content for AI training. This uncertainty is likely to have a chilling effect on GLAMs wishing to take advantage of AI technologies. This is one reason why we argue that the use of copyright works to train AI should be considered non-infringing by default.

As concerns [CC-licensed content](#), where copyright permission is required to train AI systems (in those cases where no exception would be available), the licenses grant that permission under different terms and conditions depending on the particular CC license. A [flowchart](#) helps visualise whether the licenses are triggered and if so, what conditions may apply.

[Beyond copyright](#), several obstacles to opening up, sharing and using GLAM collections related to ethics, privacy and data protection need to be assessed to bring clarity to the rapidly evolving role that AI is playing in our society and in the GLAM sector in particular. Our position takes account of the fact that other concerns beyond copyright may need to be contemplated when considering access to data to train AI.

Indeed, many issues raised by the development of AI are to be addressed under the lens of ethics, cultural rights and interests, fundamental human rights (including the principles of equality and non-discrimination), personality rights, privacy rights, and data protection. These adjacent issues should be addressed and debated in their respective policy arena, not within the framework of copyright. Consideration of the copyright implications can be undertaken while bearing in mind that these issues need to be satisfactorily addressed and resolved in their own policy sphere. Noting that those other issues are key to a coordinated and inclusive policy approach on AI and that they have a direct impact on any copyright discussions on AI, the present submission is nevertheless limited to some of the substantive questions that AI raises in the copyright arena.

Be that as it may, the objectives of allowing unfettered access to copyright-protected content to train AI include, inter alia: to reduce bias and improve inclusion; to foster innovation in AI; and to promote legitimate activities such as education and research, etc. Our position comes within the scope of supporting all Canadians' fundamental rights, especially freedom of expression, access to information and the principle of non-discrimination.

Regarding text and data mining (TDM), such activities are pivotal in supporting research and innovation and in the training of AI systems. TDM activities are non-consumptive and non-expressive uses of work. TDM does not compete with original markets for works, and may indeed enhance them by increasing demand for a wider range of works. TDM should not be made subject to additional authorisations or payments once access is legitimate. Generally, TDM activities should not be considered copyright infringement and should not be restricted by copyright. TDM should be allowed and supported pursuant to exceptions and limitations, in particular to enable a proper exercise of the users' rights.

In light of the above:

- we encourage the use of larger and more diverse sets of data in order to avoid bias in outputs and limit risks of discrimination.
- Provided access is legitimate at the point of input, broad and open exceptions and limitations should apply to support the most extensive possible use of copyrighted works for AI purposes in order to encourage the elimination and minimization of bias and discrimination.
- Placing barriers around copyright material that can be freely text-and-data-mined risks increasing the likelihood of AI bias, unfairness, discrimination and exclusion. One way to reduce those in AI systems, in addition to ensuring that the algorithm itself is not biased, is to ensure that the maximum volume and widest diversity of content is available for training purposes, requiring both minimising unnecessary barriers to technological protection measures (TPMs) and facilitating uses across borders.
- Freedom of expression and access to information must be upheld.

No copyright in AI-generated outputs

AI has been seen to generate content through processes such as [Markov chains and artificial neural networks](#) like GPT-3 (Generative Pre-trained Transformer 3, a deep learning model that can produce text). Such content might very well become part of GLAMs' collections as it starts to gain appreciation as a new form of "creative" expression. Likewise, the content generated by GLAMs using AI technology (like enriched datasets) is likely to become abundant as more and more institutions explore the opportunities offered by AI.

While the copyright status of such content is unclear under existing law, CC is of the firm view that there should be [no copyright on AI-generated content](#) and that it should be [in the public domain](#). Public domain material can be widely accessed, used and reused by GLAMs in fulfilment of their public-interest mission as well as by the general public. Canadians all benefit when knowledge, culture, and history are made accessible and shareable. We generally [advocate](#) for open access to knowledge and culture and resist further enclosures of our shared public domain.

Likewise, there should be no other copyright-like protection, such as neighboring or ancillary rights over AI-generated content. Using copyright to govern AI output is contradictory to copyright's primordial function of offering an enabling environment for human creativity to flourish. There can thus be no protection where there is no human author and where the generated content is not original. One would be mistaken to presume an equivalence between human-authored works and AI-generated outputs.

Furthermore, copyright is not the right mechanism to encourage economic investment in the development of AI systems. Copyright's utilitarian doctrine and incentives theory cannot support a claim that AI be afforded rights for any generated output because AI fails to meet the role of the author and its contribution to human-led social progress. Granting AI-outputs the status of copyright work goes against the social purpose for which copyright was created. In addition, such protection would in fact harm creators by increasing the liability risk of "real" authors and reducing the resources available in the public domain upon which further creativity depends.

Lastly, granting copyright to AI output would raise important concerns around intellectual property rights overlap leading to overprotection. Other exclusive rights (including copyright and patent rights in AI software) already provide sufficient protection in the AI space. Overprotection can have negative impacts on creativity, innovation and the provision of public goods. Content generated without human creative input should be in the public domain.

Technological Protection Measures

CC has long [disagreed](#) with the use of digital rights management (DRM) and technological protection measures (TPMs) in the open environment. We believe that DRM and TPMs should not be used to control, limit, prevent or otherwise affect activities and uses allowed under CC licenses' terms. Plainly, DRM and TPMs are antithetical to the "open" ethos and at odds with the values of sharing that we support.

Generally, we encourage creators to share their content in "downloadable" and "editable" formats (i.e. DRM-free — without any technical restriction to download, copy, or modify) to make it easier for others to benefit from and use the content, including for educational and socially beneficial purposes. We likewise discourage sharing CC-licensed content on platforms, sites or channels that add DRM to the shared content. That way, the spirit of open licensing is upheld

and the legitimate expectations of the public regarding the freedoms associated with using openly-licensed content are not compromised.

Standing up [against DRM](#) is incredibly important for many communities in the open movement, particularly open education. Of particular importance is the ability for educators and learners to “[retain](#)” content and “to make, own, and control copies of the content (e.g., download, duplicate, store, and manage).”

DRM poses a dire risk to the principles at the foundation of the open movement. DRM often constitutes an unnecessary obstacle preventing access to and use of content for legitimate purposes. When used in connection with openly-licensed content, DRM does a disservice to the public: it blocks legitimate access to the content, thereby posing a threat to the universal, fundamental rights of access to knowledge, science, culture and education.

In short, there should be no DRM or TPMs to restrict or prevent (otherwise legal) access to the data. There should be ethical requirements for transparency in the modalities of use of data, however these should be established outside the boundaries of the copyright system.

[End of document]