Internet Society
Canada Chapter

Submission to Innovation, Science and Economic Development Canada re: Copyright in the Age of Generative Artificial Intelligence

By

Internet Society Canada Chapter

January 15th, 2024

COPYRIGHT IN THE AGE OF GENRATIVE ARTIFICAL INTELLIGENCE

EXECUTIVE SUMMARY

1. The Internet Society Canada Chapter (ISCC) welcomes the opportunity to provide insight on how to leverage amendments to the *Copyright Act* to ensure Canadians can harness the full benefits of generative artificial intelligence (AI) and other AI technologies.

2. ISCC agrees with the objectives of the *Copyright Act* and believes that it must facilitate the creation of a positive environment for investment in AI development in Canada. Our laws should ensure that Canadian works are reflected—not excluded—from global AI systems and that Canadian creators have the means to protect their works from copyright infringement due to AI-generated outputs in appropriate cases. As AI further influences how information is gathered, searched and found online, and therefore the knowledge that shapes worldviews, Canadian data, language, values and culture must be included.

3. More specifically, the ISCC submits that:

    a. Recognition of the difference between the "inputs" and "outputs" of Large Language Models (LLMs) is imperative;
    b. A modern copyright framework for generative AI must ensure that Canadian data, language and values are reflected in the inputs to LLMs to protect Canadian culture and reflect it in world-shaping AI technologies;
    c. Canadian copyright law must be clarified to promote competition, innovation, and investments in Canadian computational power.

ABOUT THE INTERNET SOCIETY CANADA CHAPTER

4. The Internet Society Canada Chapter (ISCC) is a member-based not-for-profit that advocates for affordable, fair and secure internet access for all Canadians. ISCC engages on legal and policy issues to promote an open internet. Our focus is to bridge the digital divide along all axes to ensure that Canadians reap the socio-economic benefits the internet provides.

5. We provide Canadians with a proactive voice on all internet issues through various committees, roundtable discussions, conferences and membership meetups, where leaders and experts from governments, the private sector, civil society, academia, the technical community and end-users can discuss key issues, identify common solutions and share resources.

INTRODUCTION

6. As stated in the consultation paper, "the [Copyright] Act to promote the creation and distribution of content, to foster investment and job creation, promote just rewards for creators, and to create a thriving marketplace that offers consumers choice ad access to diverse content." These are all objectives that ISCC agrees with.

7. While Canada is a global leader in AI fundamental R&D, it continues to be a laggard with regards to commercialization and adoption. This is an issue central to Canada's ability to compete in the global digital economy. It is also critically important that Canadians have the skills needed to commercialize and scale adoption of AI-driven technology. If Canada and Canadians are to reap the economic, innovation and cultural benefits of rapidly evolving AI technologies, our copyright law must facilitate the creation of a positive environment for domestic AI investment, development and commercialization.

8. Our laws should ensure that Canadian works are not excluded from global AI systems such that they become inaccessible and irretrievable as an unintended consequence of overly restrictive Canadian copyright law.

9. Canadian copyright law should ensure that Canadian creators have the means to protect their works from copyright infringement due to AI generated outputs in appropriate cases.

10. ISCC is concerned that amendments to the *Copyright Act* could effectively place a toll or limit on the ability to use Canadian content to develop and train large language models (LLMs), ultimately harming Canadian consumers and Canadian creators. Copyright law reflects a careful balance between the rights of creators and the rights of users of creative content. Any recalibration of this balance needs to be very carefully considered to avoid unforeseen or unintended consequences.

11. Canada has generally only made changes to its copyright laws in harmony with our main trading partners. A change to our copyright laws that is out-of-step with the international community will inevitably create barriers and disadvantage Canadians. A mandatory licensing scheme will result in Canadian content being excluded from learning models, and thus not available for Canadians to use.

COPYRIGHT AND THE CREATION OF LARGE LANGUAGE MODELS

12. Any discussion about copyright and large language models (LLMs) needs to start with a common understanding of how LLMs work. The discussion should also distinguish between the "inputs" of an LLM and the "outputs".

13. To create an LLM, a large data set is ingested, and the words—or portions of words—are analyzed in terms of probabilistic distribution. Software examines the data set and encodes what words are in the data set and what words follow other words. It evaluates the context in which particular words appear. The result is a massive table of numerical tokens that represent words, their frequency, and their tendency to appear together. Other patterns in the appearance of words may be mapped. This creates a word prediction model, not a data set of the works examined or the training data itself. Thus, any "copying" of the training data is incidental, ephemeral, and temporary. The creation of an LLM should not trigger any of the exclusive rights of a holder of copyright under s.3(1) of the *Copyright Act.*

14. The output of an LLM depends on the model itself and the prompts given to the system by the user. One cannot enter a citation of an article or any other work and ask for a copy of it. The database

would not contain the work. ISCC believes that, at present, the *Copyright Act* can effectively address the outputs and requires no amendment in this regard.

15. In the event that a clearly constructed prompt results in an output that appears to be a word-for-word copy of an existing text, it is due to the probability that words exist in this particular sequence—not because the original text is stored in the LLM. For example, if you entered the following prompt into an LLM, "*finish the sentence: it was the best of times, it was …"* the result would probably be "it was the best of times, it was the worst of times," because of the number of times this phrase has been repeated on the internet, not because the LLM contained a copy of "A Tale of Two Cities."

16. Understanding the difference between inputs and outputs are imperative, as any radical changes to the *Copyright Act* because an LLM appears to be copying an artistic work would be misplaced. The use of a copyrighted work is not *prima facie* infringing. Canadian copyright law has never countenanced restricted or prohibited learning from a copyrighted work or describing a copyrighted work. Creating a new author's right out of thin air is not consistent with the existing statute or the case law with respect to fair dealing.

17. To the extent clarity is desirable, Canada should follow the leads of Japan, South Korea and Israel, all of whom have clarified that the input, the training of LLMs, is not a violation of their copyright laws. The following clarification to the *Act* is recommended:

> **29.23.1** It is not an infringement of copyright for a person to use a work or multiple works for the purpose of information analysis, including the comparison, classification or other analysis of information pertaining to language, sound, images or other elements constituting information extracted from a work, including the creation of systems and databases to support an artificial intelligence system.

PROTECTING CANADIAN CULTURE, LANGUAGE AND VALUES THROUGH CANADIAN INPUTS TO LLMS

18. While we are at the beginning of the AI revolution and just starting to see the technology's utility, it is clear that AI is an important way that people seek answers to questions. Thirty years ago, people went to a library. Currently, we use internet search engines. Those search engines are becoming enhanced by AI to understand our questions and match them to appropriate search results. The next step will the use of AI to match a particular question to an appropriate answer.

19. In effect, AI will shape our understanding of world. It's of utmost importance that Canadian data, language, values and culture are reflected in the information-seeking and world-shaping activities powered by LLMs and other generative AI technologies.

20. If barriers are created that effectively discourage or prohibit the use of Canadian works to build LLMs, the result is that Canadian data will be excluded. These LLMs, therefore, would not be able to provide answers or results that fully reflect Canada. They may produce answers "about Canada," but

only from data originating outside of Canada—about us, not by us. LLMs used in Canada, by Canadians, should include data that is relevant and appropriate for Canadians. To do so otherwise would be actively harmful.

21. Canadian cultural policy generally rests on the concern that Canadian culture and its cultural products may be overwhelmed by those from the United States; moreover, that Canada's bilingual and multicultural legacy may be diluted. Any scenario in which there is mandatory licensing or other restrictions on the use of Canadian content in the creation of LLMs will obscure results about Canada to Canadians and the rest of the world. This would be a bad outcome.

CANADIAN COPYRIGHT LAW MUST BE CLARIFIED TO PROMOTE COMPETITION, INNOVATION, AND INVESTMENTS IN CANADIAN COMPUTATIONAL POWER

22. The organizations that fund, build and operate large-scale computing facilities are closely examining what the next generation of computing will be, and are sensibly examining the regulatory environment to evaluate long-term investments in computing capacity.

23. Canadians benefit from having computing and networking infrastructure built domestically. It creates jobs, boosts skill sets across the entire economy and creates greater capacity to access information. Moreover, domestic computing and networking infrastructure is important for data sovereignty and national security, as is recognized in other government initiatives, such as Bill C-26, *An Act Respecting Cyber Security*.

24. Right now, AI is a significant driver for increasing computing capacity. AI requires specialized chips and new technologies that are built-for-purpose. Before making decisions about where to deploy expensive computing capacity, companies will look to where and whether they and their customers can use it in the location of deployment. Any changes to the law that will increase the legal or regulatory risk associated with the use of infrastructure will disincentivize investment in computing capacity in Canada, to the detriment of the socio-economic benefits for Canadians.

25. Simply put, Investment in computing capacity will be influenced by the degree of clarity provided by copyright law for potential investors. Canada has an opportunity to glean valuable insights from the experiences of its international partners—for example, where clarity arising from text and data mining (TDM) exceptions have succeeded in other jurisdictions.

26. In Canada, a clear exception for TDM is crucial to encourage competition and innovation from smaller, Canadian players. If copyright law requires licensing of the internet—or the corners of it originating in Canada—before an LLM can be trained, then only the biggest players with the deepest pockets and hoards of proprietary data may be able to innovate. This would be a massive barrier to entry for an upstart company and would prohibit Canadian companies with Canada's best interest at heart from competing and innovating.

CONCLUSION

27. The questions posed in the consultation paper will have wide-ranging effects on Canadian culture and the economy. Decisions made as a result of the consultation will have a significant effect on whether and how Canadians can access information created by Canadians, for Canadians, about Canada. It will be consequential for long-term investments in Canadian computing capacity, which Canadians should be able to benefit from.

28. ISCC hopes that our perspective has been helpful and welcome any opportunities to further discuss this important topic.