

September 4, 2018

Submission of Microsoft to the Standing Committee on Industry, Science and Technology on the 2018 Statutory Review of the Copyright Act

Microsoft welcomes the opportunity to participate in the Standing Committee on Industry, Science and Technology's review of the Copyright Act. This review comes at a critical moment when individuals, industries, governments, and researchers are increasingly implementing digital innovations, such as artificial intelligence, as part of a surge of digital transformation.

Our submission highlights the importance of promoting the benefits of this transformation by ensuring that Canadian copyright law does not inhibit the development of artificial intelligence (AI) and Canada's modern digital economy, and preserves the right to read, understand, and analyze information – activity that has never been subject to control under copyright laws anywhere.

Towards that goal, it is critical that the Government of Canada encourage and foster an environment where AI is not burdened by unnecessary copyright regulation and remains accessible to all entities and individuals for all purposes. To do this, Canada should join the many countries that have unequivocally established sufficient protections to ensure that AI-related activities, such as machine learning, are widely adopted and used to their full potential.

Promoting Knowledge Innovation

How we develop knowledge is changing dramatically. We are awash in information today, and digital transformation requires new ways to read, analyze and understand the vast array of information. Artificial intelligence is critical to that effort, and machine learning forms the backbone for AI. Machine learning relies on aggregating both raw and structured data and content into a machine accessible form, and analyzing this information – at hyper scale – to identify insights, patterns and relationships.

By using machine learning to analyze the information in millions of diverse works, computers are now robust enough to model human perception and help augment how we make critical decisions. The output of this process is mathematical models that enable machines to make predictions and take actions without needing to be explicitly programmed.

The impact of machine learning on daily life can be life changing. As an example, researcher Cecily Morrison and a team of developers at Microsoft's research facility developed a unique application for computer vision that uses artificial intelligence build from machine learning.

“*Seeing AI*” lets sight-impaired persons understand who and what is around them.¹ It does so by examining visual content and extracting and identifying visible objects, as well as information about those objects – the age and gender of a depicted person, whether they are smiling or frowning, what they are doing. To accomplish this, *Seeing AI* utilizes millions of publicly available images depicting various objects – trees, sign posts, animals, landscapes, business interiors – to improve the ability of machines to recognize and inform users about the environment around them.

Machine Learning offer significant benefits to everyone

The ability to harness the benefits of AI exists for everyone – large or small, public or private, commercial or non-commercial. All can readily use AI to develop innovative projects benefitting the public.

Start-ups, research groups, academics, not-for-profits, government and businesses are increasingly using machine learning, aided by processes like text and data mining, to develop algorithms to learn from data and to understand business trends, research new markets, and develop new technologies and applications.²

The societal benefits of broad machine learning are readily apparent. Examples include:

- identifying if and how CRTC decisions are based on the public’s input;³
- predicting disease outbreaks by analyzing online news media and other data;⁴
- developing individualized cancer treatments by analyzing the latest information about effective treatments;⁵
- assessing specific community well-being by crawling and analyzing online hyperlocal posts and open data sources;⁶
- identifying foodborne disease risks by harvesting information buried in publications, government reports, datasets and social media;⁷

¹ <https://www.microsoft.com/en-us/seeing-ai> and https://www.youtube.com/watch?v=bqeQByqf_f8

² <https://www.crunchbase.com/category/machine-learning/5ea0cdb7c9a647fc50f8c9b0fac04863> (listing over 1000 businesses, start-ups, and projects using machine learning).

³ <https://www.cybera.ca/news-and-events/news/new-canadian-data-mining-tool/>

⁴ <https://www.technologyreview.com/s/510191/software-predicts-tomorrows-news-by-analyzing-todays-and-yesterdays/>

⁵ <https://news.microsoft.com/stories/computingcancer/>

⁶ <http://escience.washington.edu/research-project/assessing-community-well-being-through-open-data-and-social-media/>

⁷ <http://www.who.int/mediacentre/news/releases/2015/food-safety/en/>

- identifying accessible sidewalk routes for impaired mobility travelers by crawling online maps and online posts;⁸ and
- identifying and addressing the proliferation of fake news.⁹

Canada's leadership position in AI depends upon broad access to data, including for commercial development of machine learning and AI

Canada has emerged as a leader in AI.¹⁰ A Canadian trio of deep learning pioneers— Geoffrey Hinton, Yoshua Bengio, and Yann LeCun— are credited with placing Canada at the forefront of AI globally¹¹, with Canada recently being recognized as having the world's third largest AI talent pool.¹²

The Government of Canada has recognized the importance of AI to the economy and the opportunity to translate Canada's leadership position in AI into jobs and economic activity. For example, the Government has asked the Canadian Institute for Advanced Research (CIFAR) to develop a \$125-million Pan-Canadian AI strategy focused on attracting, retaining and training AI research in the country.¹³ As well, it has announced \$950-million of funding for innovation superclusters, including the Digital Technology Supercluster¹⁴ and the AI-Powered Supply Chains Supercluster,¹⁵ each of which will involve analyzing large datasets to advance machine learning and AI solutions.

Canada is not alone in its ambition to be a powerhouse in artificial intelligence. A "new tech arms race"¹⁶ has emerged among many of the world's leading economies, including US, Canada, the UK, China and Japan.

As the competition among these countries for human talent, financial capital, and commercialization opportunities intensifies, it is clear that access to data is a critical competitive variable. To train machines that can recognize objects, speak, listen, interpret, and more, it is necessary to feed computers huge amounts of data so that they can look for

⁸ <http://escience.washington.edu/research-project/sidewalk-maps-for-low-mobility-citizens/>

⁹ https://www.theparliamentmagazine.eu/articles/partner_article/apco-worldwide/unleashing-big-data%E2%80%99s-potential-journalism-economy-and

¹⁰ <https://www.universityaffairs.ca/features/feature-article/canada-emerged-leader-artificial-intelligence/>

¹¹ <http://www.canadianbusiness.com/innovation/rbc-brain-drain-deep-learning/>

¹² <https://www.macleans.ca/technology-3/four-ways-canada-can-own-the-artificial-intelligence-century/>

¹³ <https://www.cifar.ca/ai/pan-canadian-artificial-intelligence-strategy>

¹⁴ <https://www.ic.gc.ca/eic/site/093.nsf/eng/00011.html>

¹⁵ <https://www.ic.gc.ca/eic/site/093.nsf/eng/00009.html>

¹⁶ <https://enterpriseiotinsights.com/20180410/channels/fundamentals/review-of-national-ai-strategies-tag40-tag99>

patterns, relationships, and insights. This reality makes it reasonable to predict a migration of talent and capital to countries with AI-friendly laws.

Some of Canada’s key competitors have extended legal protections for broad Machine Learning

Outside of Canada, countries, including Japan, the United States and China, have extended and are expanding legal protections for broad machine learning techniques including text and data mining. Japan recently implemented changes to its copyright laws that significantly expanded and clarified an already forward-looking machine learning exception.¹⁷ The UK permits text and data mining for certain purposes, and is currently exploring broadening its exception as it increasingly recognizes the benefits of machine learning for all users. China, Singapore and Thailand are focused on similar digital copyright reform and have similarly proposed broad and unrestricted machine learning exceptions.¹⁸ Unsurprisingly, these countries are also at the forefront of research involving data analytics and artificial intelligence, including machine learning.¹⁹

In contrast, countries that have proposed a more limited approach²⁰ have received strong opposition by leading research²¹ and start-up organizations²², who understand that research in the 21st century occurs in all areas of industry, and frequently involves fluid collaborations between public and private entities. Mindful of preserving a role in the next great technical revolution and stopping the “brain drain” of researchers to countries with friendlier AI-policies, top European universities in AI research have stressed support for the strong public-private collaborations that form the backbone for AI research, and have urged the European Commission to adopt friendlier policies that favor the commercial development of machine learning and AI.²³ The EU Parliament shared these concerns, as it recently voted to revise a

¹⁷ See http://www.cric.or.jp/english/clj/cl2.html#cl2_1+SS5 for the new Article 30-4 (allowing broad text and data mining), Article 47-4 (enabling incidental reproductions) and Article 47-5 (allowing creation and storage of searchable datasets of copyrighted works).

¹⁸ See Proposal 9: Unlocking the potential of text and data mining, at <https://www.mlaw.gov.sg/content/minlaw/en/news/public-consultations/public-consultation-on-proposed-changes-to-copyright-regime-in-s.html>

¹⁹ <https://www.whitehouse.gov/sites/whitehouse.gov/files/images/EMBARGOED%20AI%20Economy%20Report.pdf>

²⁰ For example, the European Commission proposed a very limited exception only for specifically defined public interest or non-commercial research organizations, a proposal that has been strongly criticized (<https://euobserver.com/opinion/135474>).

²¹ <http://libereurope.eu/wp-content/uploads/2015/11/TDM-Copyright-Exception.pdf> (“There is no reason to limit the [ML] solution to non-commercial uses...”)

²² <https://www.euractiv.com/section/digital/opinion/the-eu-just-told-data-mining-startups-to-take-their-business-elsewhere/>

²³ See ELLIS’s “Open Letter to the European Commission”, <https://ellis-open-letter.eu/letter.pdf>

copyright proposal that would have significantly impacted the use of text and data mining by startups and businesses of all sizes.²⁴

Copyright should not impede our ability to understand, analyze, and learn

Just as copyright has never controlled how people understand, research, or analyze the books they read, copyright should not impede the use of technology to enhance the perception and analysis by machines of lawfully accessed works. With machine learning, it is often necessary to make copies of lawfully acquired information. These copies are not read by humans. Nor are they consumed or redistributed for their creative expression. They do not substitute for or displace the markets for original articles or subscriptions.²⁵

Most researchers and innovators would not expect that their machine learning projects involving lawful access to works would be impeded by copyright. However, merely because machine learning techniques, such as text and data mining, may require incidental copying of works, doubt or uncertainty around the legality of various machine learning techniques may exist under applicable copyright regimes.²⁶

Providing clarity around machine learning, and ensuring that its techniques are available to all, removes legal ambiguity around machine learning and helps to unlock the potential for innovative research by both public and private sector. It also encourages start-ups, businesses, and private researchers to embrace AI projects, both independently and via public-private collaborations.²⁷

Why clarity is necessary

Like other copyright laws around the world, Canadian copyright law does not preclude the ability of users to read, analyze, and understand lawfully acquired works. It does not, however, include explicit references that clarify that the scope of copyright does not extend to activities undertaken for this purpose.

Because machine learning techniques may require the incidental copying of lawfully acquired copyrighted works to make them accessible for machine learning, analyze them for patterns, facts, and insights, and use those copies for data verification, there is a risk that copyright can be asserted as a basis to block this activity unless permission of a copyright owner is granted.

²⁴ See (<https://sciencebusiness.net/news/parliament-vote-sends-controversial-copyright-law-back-drawing-board>)

²⁵ See UK Government, *Modernising Copyright: A Modern, Robust and Flexible Framework* (2012), 37. "...the copying involved in text and data analytics is a necessary part of a technical process and is unlikely to substitute for the work in question (such as a journal article)."

²⁶ See e.g. <https://www.alrc.gov.au/publications/8-non-consumptive-use/text-and-data-mining>.

²⁷ http://www.lisboncouncil.net//index.php?option=com_downloads&id=1262 ("Given ever closer partnerships and collaborations between publicly funded research institutions and companies....we believe that the only workable and justifiable solution is the least ambiguous one...")

While the *Copyright Act* already provides several exceptions that permit the use of copyright materials without the need to obtain prior permission,²⁸ none of these exceptions explicitly address machine learning activities. Further, while certain fair dealing exceptions may apply to some machine learning activities,²⁹ it is at best uncertain as to whether these exceptions will be interpreted to allow for the broad access necessary for all AI users in all sectors. As a result, both the many societal benefits of AI and the commercialization of AI – which is critical to the Government of Canada’s plan to translate Canada’s leadership position in AI into jobs and economic activity – are at risk of being stymied at the very moment they are poised to deliver meaningful impact.

What needs to happen

For Canada to affirm its position as a leader in AI it needs to eliminate uncertainty and doubt about the legality of machine learning techniques for all users. This can be accomplished by adding express exceptions to the *Copyright Act* to permit the use of lawfully acquired copyrighted works in machine learning techniques by any entity for any lawful purpose. This solution should be drafted to clarify that the steps of machine learning – copying lawfully acquired works to analyze and develop new knowledge, and verify such findings – are permitted without authorization of the copyright owner, and by any entity or individual for lawful purposes. Reference should be made in the legislative record to: (i) clarify that the scope of copyright law does not extend to machine learning, and (ii) recognize the importance of machine learning and artificial intelligence and preserving access to unprotected elements of copyrighted works.

How potential concerns of content owners can be addressed

Copyright owners whose works are lawfully acquired by machine learning users are not harmed by any approach that clarifies that copyright cannot be utilized to restrict machine learning techniques, as nothing under applicable copyright law limits their use of non-copyright measures to restrict access. However, for users that have lawfully acquired material protected by copyright, copyright owners should not be able to use copyright law to restrict or create uncertainty about techniques used for machine learning.

²⁸ For example: reproduction for an individual’s private purposes (section 29.22), backup copies (section 29.24), and reproduction for the purpose of encryption research (section 30.62).

²⁹ For example, the exceptions in respect of research, education and private study (section 29).

Summary of Recommendations

- Add express clarification that the *Copyright Act* does not prohibit machine learning by any entity for any lawful purpose.
- Draft the new exceptions to clarify that the techniques of machine learning involving copying analyzing, and using lawfully acquired works protected by copyright to develop new knowledge – are permitted and require no authorization of the copyright owner, and by any entity or individual for lawful purposes.

About Microsoft

Microsoft (Nasdaq “MSFT” @microsoft) enables digital transformation for the era of an intelligent cloud and an intelligent edge. Its mission is to empower every person and every organization on the planet to achieve more.