

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

Professor Pina D'Agostino

10 December 2018

Introduction

I am Professor Pina D'Agostino, an Associate Professor at Osgoode Hall Law School of York University and the Founder and Director of IP Osgoode, the law school's flagship Intellectual Property Law & Technology Program. Launched in October 2008, IP Osgoode cultivates interdisciplinary, comparative and transnational research, collaboration, policy-thinking, and practice on intellectual property ("IP"). IP Osgoode is an innovative program that is cutting-edge and unique in Canada and has global recognition.

We aim to provide balanced and objective research, offer new and unexplored viewpoints to public policy discussions, which are inclusive of the opinions and interests of a broad range of IP stakeholders (including governments, NGOs, the legal community, businesses and the general public), and ultimately, act as a facilitator for the flourishing of a knowledge-based society in Canada.

Earlier this year, IP Osgoode, along with our collaborators, organized a highly successful round table and conference focused on legal and ethical issues related to artificial intelligence ("AI") (the proceedings of the conference are available at <http://aichallenge.osgoode.yorku.ca>). We aim to hold more such events and activities on the issues that emerged from the AI roundtable and conference, facilitating open discussions about AI, Big Data and other emerging technologies.

I am writing in my personal capacity for the purpose of this submission.

Recommendation

I believe that *sui generis* legislation for AI may be necessary in the future. However, for the statutory review, I recommend the introduction of a *Text and Data Mining (TDM) exception* into the *Copyright Act*. The exception should be a separate clause in the *Copyright Act* rather than part of the existing fair dealing framework.

At this time, a TDM exception is needed to facilitate the successful dissemination of useful data and information across society and to promote the commercialization of AI technology. Such an exception aligns with the Government of Canada's National IP and Innovation Strategies as well as federal investments in AI across the country. TDM exceptions have been successfully adopted outside of Canada.

Background

The Government of Canada has made development and commercialization of IP a priority in its IP strategy.¹ One opportunity in technology commercialization today is AI.² AI systems do not fit easily within other technological categories, as AI-enabled technologies process data and generate products by identifying similarities and patterns in the data that even its programmers and operators may be unaware of. As such, AI has the enormous potential to make almost every industry more efficient through better management of information. The Government of Canada is investing in AI as a means to foster socio-economic benefits for Canadians.³ Similarly, companies across the country in a wide variety of industries are currently investing in AI technology.⁴ Canadian copyright law will play a crucial role in fostering both public and private goals.

¹ Canada, *Intellectual Property Strategy*, Innovation Science and Economic Development Canada, online: <www.ic.gc.ca/eic/site/108.nsf/eng/home>. The IP Strategy states that its purpose is to *help extract value from existing products and services, create new revenue streams and raise capital*.

² See e.g. Shlomit Yanisky-Ravid & Xiaoqiong (Jackie) Liu, "When Artificial Intelligence Systems Produce Inventions: The 3A Era and an Alternative Model for Patent Law" (2017) 39 *Cardozo L Rev* 2215.

³ Canada, *Growing Canada's Advantage in Artificial Intelligence*, Department of Finance Canada (30 March 2017), online: <https://www.fin.gc.ca/n17/17-026-eng.asp>; CIFAR, *Pan-Canadian Artificial Intelligence Strategy*, Canadian Institute for Advanced Research, online: <https://www.cifar.ca/ai/pan-canadian-artificial-intelligence-strategy>.

⁴ See e.g. Boeing's recent investment in AI: "Boeing Invests in Robotics Institute Autonomous Flight Spin-Off" (19 October 2017), online: *Carnegie Mellon University News* <www.cmu.edu/news/stories/archives/2017/october/boeing-investment.html>; See also "The Race for AI: Google, Intel, Apple in a Rush to Grab Artificial Intelligence Startups" (27 February 2018), online: *CB Insights* <www.cbinsights.com/research/top-acquirers-ai-startups-ma-timeline/>.

Developing AI systems requires the input of enormous amounts of information into a computer to “teach” it how to think, a process known as machine learning (“ML”). In order to proceed with ML, AI developers need to extract data and information from a wide variety of sources through processes such as TDM.⁵ Therefore, providing these AI developers access to high quality data is important. Low quality data – which is usually free and accessible – may create bias in AI systems and contribute to harmful effects.⁶

Unfortunately, copyright law in Canada presents serious obstacles to persons seeking to use high quality data. Our current regime grants exclusive rights of copying to copyright owners, so that a developer seeking data for AI/ML training purposes must obtain licences for each individual copyright-protected work.

Since the development of AI systems requires millions of pieces of data, and statutory damages for copyright infringement can run to \$20,000 *per each unauthorized copy*,⁷ there is currently an enormous legal and financial risk to conducting AI research, which could deter both small and big tech companies from investing or operating in Canada.

A TDM Exception is Important for Canada’s IP & Innovation Strategies

If Canada wishes to foster innovative AI research, it is crucial that high quality data be readily accessible to developers. Because of the nature of ML, AI is only as good as the data used to train it. When an algorithm is trained on data fraught with gender or racial biases, for instance, it will produce predictions and products that perpetuate or even magnify the same biases.

Large Internet platforms, such as Google and Facebook, have terms of service that typically grant the owners broad licences for any IP uploaded by users, and typically enjoy access to high-quality data. On the other hand, many smaller players lack this data access.

With no reasonable route to access data, Canadian AI developers will opt for low-risk data that are easily obtained but also typically of poor quality. The law also encourages

⁵ TDM can be defined as “the discovery by computer of new, previously unknown information, by automatically extracting and relating information from different resources, to reveal otherwise hidden meanings.” See Future TDM, “What Is Text and Data Mining?” (23 February 2016), online: *The Future of Text and Data Mining* <www.futuretdm.eu/news/tdm-definition/>.

⁶ E.g., Ryan Calo, “Artificial Intelligence Policy: A Primer and a Road Map” (2017) 51:2 UC Davis L Rev 399; Amanda Levendowski, “How Copyright Law Can Fix Artificial Intelligence’s Implicit Bias Problem” (2018) 93 Wash L Rev 579.

⁷ *Copyright Act*, RSC 1985, c C-42, s 38.1(1).

corporations to create “black box” systems that obfuscate the source of their data, and thus reduce accountability for copyright infringement and for the use of biased algorithms.

Reforming Fair Dealing or Introducing Fair Use is the Wrong Approach

Some propose that Canada’s fair dealing provision should be amended to add a new purpose of TDM. Such proponents draw support from the Supreme Court of Canada (SCC) decisions since *Théberge*,⁸ which have upheld fair dealing as a user right to be interpreted in a broad and liberal manner.⁹ Others point out that even under this approach, a permitted purpose might fall outside of fair dealing, since many AI outputs will be for commercial purposes, a factor that may often negate fair dealing.¹⁰

These limitations have led others to argue that Canada should adopt the broader US doctrine of fair use. In this model, they suggest a court would look at how a work is used, as a research activity, rather than the purpose of the use, which would be better suited to AI technology and other digital uses of works.¹¹

Introducing fair use into Canada’s *Copyright Act* will only substitute the uncertainty of fair use for the uncertainty of fair dealing: courts will be faced with an interpretation fog, completely lacking precedents or having somehow to adapt foreign precedents to Canadian circumstances.¹²

Even in the US, scholars have long criticized US fair use, and have specifically argued against adopting the fair use model for AI.¹³ To rely in Canada, on either fair dealing or fair use to address TDM will therefore leave parties unsure on what they may lawfully do, and ultimately unsure how courts will adjudicate any disputed use.¹⁴

⁸ *Théberge v Galerie d'Art du Petit Champlain Inc*, 2002 SCC 34 [*Théberge*].

⁹ Michael Geist, “Why Copyright Law Poses a Barrier to Canada’s Artificial Intelligence Ambitions” (18 May 2017), online (blog): *Michael Geist* <www.michaelgeist.ca/2017/05/copyright-law-poses-barrier-canadas-artificial-intelligence-ambitions>.

¹⁰ *Ibid.*

¹¹ *Ibid.*

¹² See for example: Barry Sookman & Dan Glover, “Why Canada Should Not Adopt Fair Use: A Joint Submission to the Copyright Consultations” (November 2009) 2:2 *Osgoode Hall L Rev* LP 139.

¹³ Benjamin L.W. Sobel, “Artificial Intelligence’s Fair Use Crisis” (2017) 41 *Colum J L & Arts* 45. Since fair use and fair dealing are open-ended exceptions, the judiciary has great flexibility in interpreting their extent. Even a single ruling limiting the scope of fair dealing for AI research would have a chilling effect on AI innovation.

¹⁴ See Giuseppina D’Agostino, “The Arithmetic of Fair Dealing at the Supreme Court of Canada” in Michael Geist, ed, *The Copyright Pentalogy: How the Supreme Court of Canada Shook the Foundations of Canadian Copyright Law* (Ottawa: University of Ottawa Press, 2013).

Uncertainty costs, and such costs will deter investment in and development of high-quality AI/ML systems. This result would be ironic, to say the least, as among the goals to intervene in the *Copyright Act* should be to craft a user-friendly, predictable environment for all stakeholders.

A Specific TDM Exception for AI is the Best Approach

Canada's cornerstone copyright law principle of balancing the various stakeholder interests—creators, owners, users and the public—should be paramount in crafting a solution. Any exception for TDM must therefore be careful to focus on legitimate purposes related to the ML/AI research that is carried out. TDM exceptions in other countries are often limited to non-commercial purposes or purposes such as “scientific research” or “innovation.”

My recommendation is that the exception be worded to allow TDM for ML/AI research and development, which may be for commercial purposes. The proposed exception would apply to any published or unpublished copyright-protected works, to which a user has lawful access. Such works may be, and need not be limited to, sound, voice recordings, film/videos, data and text, images, tables and databases.

The proposed exception would allow data to be mined for AI training purposes, for instance, such as creating a database for machine learning algorithms. This exception would not allow using data that is not lawfully accessible.

In this way, Canadian stakeholders can have the clarity they need and be more at a level playing field in accessing a wide array of data within the parameters of the law. The suggested exception, which is broader than its UK counterpart for instance, importantly allows for commercial uses of data so as to incent Canadian research and development across any sector, small and large.

Ultimately, the proposed exception would complement the Canadian government's significant investments in elevating AI-driven social benefits and in developing a globally competitive Canadian-made AI industry.