



June 4, 2018

The Director, Copyright Law Section
Department of Communications and the Arts
GPO Box 2154
Canberra ACT 2601

Re: Copyright Modernisation Consultation

Dear Director

Microsoft Corporation welcomes the opportunity to respond to the Australian Government's call for views regarding Copyright modernisation. We address Question 1 of the consultation (applicable to exceptions to the Copyright Act of 1968) and focus specifically on the need for an exception permitting text and data mining (TDM) of copyrighted works.¹

All around the world, governments, businesses, start-ups, research organizations, universities, and even individuals are investing in Artificial Intelligence (AI) to help solve complicated problems across a wide range of fields, including safety, transportation, education, health, agriculture, and medicine.² As we discuss below, TDM is essential to both the development of AI and to our modern digital economy - it is critical that governments encourage and foster an environment where TDM is not burdened by unnecessary copyright regulation and remains accessible to all entities and individuals for all purposes.

Whether Australia utilizes an express exception or a broader fair use standard to ensure the availability of TDM, it should join the many countries that have unequivocally carved out sufficient protections to ensure that TDM is widely adopted and used to its full potential.

TDM Is the Foundation for AI

Due to widely available cloud computing power, more advanced algorithms, and the massive growth in machine addressable data, AI's use has exploded in recent years. Microsoft believes

¹ TDM in this submission refers to the analysis of lawfully acquired material that may be subject to copyright protection for better understanding of the facts, concepts and ideas contained in that material. We are not addressing any privacy or other regulation of the use of personal data here.

² See e.g. <http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation> (evaluating the global scope and impact of data analytics).



that everyone deserves the ability to harness these breakthroughs and enjoy their benefits both in their work and personal lives.

TDM plays a foundational role in the improvement of AI. To train machines that can recognize objects, speak, listen, interpret, and more, computers are fed massive amounts of data via the process of machine learning. Typically, computers analyze this material to look for patterns, relationships, and insights, or even just to improve their capability to understand human speech, recognize objects, and develop insights that can only be gained by churning through millions of diverse works to create algorithms robust enough to model human perception and decision making.³ The output of this process is mathematical models that enable machines to make predictions and take various actions without needing to be explicitly programmed.

As an example of the power of AI and TDM to impact daily life, researcher Cecily Morrison and a team of developers at Microsoft's research facility developed a unique application for computer vision that uses artificial intelligence built from text and data mining. "Seeing AI" lets sight-impaired persons understand who and what is around them.⁴ *Seeing AI* examines visual content, extracts and identifies 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 – the kind of things that sight enabled humans can do naturally and easily but which the sight-impaired struggle with. 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.

TDM Drives Advances in Data Analytics

In today's digital economy, TDM has emerged as powerful tool for entities in every sector.⁵ These include startups and large businesses looking to understand the impact of new products, governments needing to assess voter reaction to policy proposals, or even the local pub needing to understand its customers' changing meal and entertainment habits - all have the capability to use TDM to innovate.⁶

The democratization of technology is evident in the rapid adoption of data analytic tools, machine learning, and artificial intelligence. Our customers around the globe are using AI and

³ See e.g. <http://www.futuretdm.eu/news/tdm-process/>

⁴ See the Seeing AI project example at <https://www.youtube.com/watch?v=R2mC-NUAmMk>

⁵ See e.g. "The Growing Importance of TDM" at <http://www.futuretdm.eu/news/tdm-landscape/>

⁶ See *Int. J. Electronic Business*, November 2013 at <https://www.sciencedaily.com/releases/2013/11/131128103831.htm>



TDM to improve customer service⁷, react to customer complaints⁸, and improve its energy and water usage⁹ among other innovations.

The societal benefits of allowing broad text and data mining are readily apparent.¹⁰ By example, oncologists are using “machine reading” AI to analyze the latest information about effective cancer treatments to develop individualized treatments. This technology, developed by Hoifung Poon, a researcher at Microsoft’s research lab, allows cancer researchers to provide customized treatments at a scale that before was not humanly possible.¹¹

TDM Offers Significant Benefits for Everyone

The ability to harness the benefits of TDM exists for everyone – large or small, public or private, commercial or non-commercial – all can readily use TDM to develop innovative projects benefitting the public. Startups, research groups, and businesses of all kinds are increasingly using machine learning, aided by text and data mining, to develop algorithms to learn from and predict data to understand business trends, research new markets, and develop new technologies such as artificial intelligence applications.¹²

Examples include:

- predicting disease outbreaks by analyzing online news media and other data,¹³
- assessing specific community well-being by crawling and analyzing online hyperlocal posts and open data sources,¹⁴
- Identifying foodborne disease risks by harvesting information buried in publications, government reports, datasets and social media¹⁵, and

⁷ Telefonica created AURA using MS AI services to better serve its customers

<https://customers.microsoft.com/en-us/story/telefonica-media-telco-cognitive-services-azure>

⁸ Liebherr used data analytics and TDM to quickly react to customer concerns

<https://customers.microsoft.com/en-US/story/the-liebherr-group-harnesses-social-media-to-stay-ahead>

Carnival Maritime uses data analytics and TDM to optimize water consumption and monitor health for its ships. <https://customers.microsoft.com/en-US/story/carnivalmaritime>⁹

¹⁰ The Value and Benefits of Text Mining. JISC, 2012. <http://www.jisc.ac.uk/sites/default/files/value-text-mining.pdf>

¹¹ See <https://news.microsoft.com/stories/computingcancer/>

¹² See <https://www.crunchbase.com/category/machine-learning/5ea0cdb7c9a647fc50f8c9b0fac04863> (listing over 1000 businesses, start-ups, and projects using machine learning.

¹³ See <https://www.technologyreview.com/s/510191/software-predicts-tomorrows-news-by-analyzing-todays-and-yesterdays/>

¹⁴ See <http://escience.washington.edu/research-project/assessing-community-well-being-through-open-data-and-social-media/>

¹⁵ See <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.¹⁶
- Identifying and addressing proliferation of fake news¹⁷

TDM Enables Beneficial Public Private Collaborations

Governments are leading the effort to harness AI technologies by working with a range of innovators, and TDM plays an important role in these public private collaborations. By example, the Tasmanian Government and oyster farmers in fourteen of the State's estuaries are deploying new system that uses in-estuary sensors, cloud computing and machine learning to improve outcomes for oyster farmers while protecting the delicate oyster ecosystem.

Innovators such as The Yield work with oyster farmers to feed data from -estuary sensors into the cloud, alongside national weather data. TDM is used to analyze data sets including weather information, tidal information, water temperature, rainfall and salinity. Environmental data and real-time sensor data allows computers to produce more accurate information allows oyster growers and regulators to make faster decisions using local information. As a result, oyster farms are only closed when they have to be, producing savings, reducing costs and boosting efficiencies that help better regulate this delicate ecosystem.¹⁸

Copyright Should Not Impede TDM

There is very little connection between copyright and TDM, just as copyright has never controlled how people read books and do research. With TDM, it may be necessary to make copies of information to train the artificial intelligence and allow it to analyze this material to look for patterns, relationships, and insights. These copies are not read by humans, nor are they consumed or redistributed for their creative expression, so they don't substitute for the original articles or subscriptions.¹⁹

Most researchers and innovators wouldn't expect that their TDM projects involving lawful access to works would be impeded by copyright. However, as Australia's government studies have demonstrated, merely because TDM may require incidental copying of works, doubt or

¹⁶ See <http://escience.washington.edu/research-project/sidewalk-maps-for-low-mobility-citizens/>

¹⁷ See https://www.theparliamentmagazine.eu/articles/partner_article/apco-worldwide/unleashing-big-data's-potential-journalism-economy-and

¹⁸ See <https://customers.microsoft.com/en-au/story/the-yield-agriculture-azure-en>

¹⁹ 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)."



uncertainty around TDM's legality may exist under applicable copyright regimes.²⁰ A broad TDM exemption, available to all, removes legal ambiguity around TDM and helps to unlock the potential for innovative research by both public and private sector, and encourages startups, businesses, and private researchers to embrace text and data mining projects, both independently and via public-private collaborations.²¹ And because lawful access to copyrighted materials is required, TDM has no impact on the business models of producers of that content.

Outside of Australia, countries have extended and are expanding legal protections for broad TDM activities, such as Japan, the United States and China. The UK and Canada also permit TDM for certain purposes but are currently exploring broadening their exceptions as they increasingly realize the impact of TDM on all users. China, Singapore and Thailand are focused on similar digital copyright reform and have similarly proposed broad and unrestricted text and data mining exceptions.²² Unsurprisingly, these countries are also at the forefront of research involving data analytics, artificial intelligence, and machine learning.²³

In contrast, countries that have proposed a more limited approach to TDM²⁴ have received strong opposition by leading research²⁵ and start up organizations²⁶, 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 TDM 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

²⁰ See e.g. <https://www.alrc.gov.au/publications/8-non-consumptive-use/text-and-data-mining>.

²¹ 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 [TDM] solution is the least ambiguous one...")

²² 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>

²³ https://www.whitehouse.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

²⁴ For example, the European Union is considering a very limited exception only for specifically defined public interest or non-commercial research organizations.

²⁵ <http://libereurope.eu/wp-content/uploads/2015/11/TDM-Copyright-Exception.pdf> ("There is no reason to limit the [TDM] solution to non-commercial uses...")

²⁶ <https://www.euractiv.com/section/digital/opinion/the-eu-just-told-data-mining-startups-to-take-their-business-elsewhere/>



the European Commission to adopt friendlier policies that favor the commercial development of machine learning and AI.²⁷

Australia can implement adequate protections for TDM in a variety of ways. An express exception to permit TDM by any entity for any lawful purpose helps to eliminate ambiguity about the various aspects of TDM that are protected. An exception should clarify that the steps of TDM –copying lawfully acquired works, analyzing the copies, and using these copies for data validation, are permitted without authorization of the copyright owner, and by any entity or individual for lawful purposes. Another approach that would be more “future-proof” would be to adopt a fair use standard (with TDM as an explicitly enumerated example) that would permit courts to reach similar results and accommodate changes in technology and economics as they develop over time. In either case, reference should be made in the legislative record to the importance of TDM and preserving access to unprotected elements of copyrighted works, and other legal precedent that has deemed such activity fair.²⁸

We welcome the opportunity to further discuss TDM and we encourage the Australian legislature to expressly preserve the many benefits of TDM available to all Australians.

Yours sincerely

Belinda Dennett
Corporate Affairs Director
Microsoft Australia

²⁷ See ELLIS’s “Open Letter to the European Commission”, <https://ellis-open-letter.eu/letter.pdf>

²⁸ See *Authors Guild Inc. v. Hathitrust*, 744 F.3d 87 (2015, 2nd Cir.); *Authors Guild v. Google, Inc.*, 804 F.3d 202 (2015, 2nd Cir.), *Fox News Network LLC. V. TVEyes Inc.*, 124 F. Supp. 3d 325 (2014) (use of news content to create a searchable archival database fair); (TDM and search function not challenged on appeal in *Fox News Network, LLC v. TVEyes, Inc.*, No. 15-3885 (2d Cir. 2018))