Australian Digital Alliance

Submission on the World Intellectual Property Organisation draft issues paper on the Impact of Artificial Intelligence on IP Policy

February 2020

The Australian Digital Alliance welcomes the opportunity to comment on the World Intellectual Property Organisation's (WIPO) draft issues paper related to the impact of artificial intelligence (AI) on intellectual property policy.

The ADA is a non-profit coalition of public and private sector groups formed to provide an effective voice for the public interest perspective in copyright policy. It was founded following a meeting of interested parties in July 1998, with its first patron being retired Chief Justice Sir Anthony Mason AC KBE QC. Its members include universities, schools, disability groups, libraries, archives, galleries, museums, research organisations, technology companies and individuals. The ADA unites those who seek copyright laws that both provide reasonable incentives for creators and support the wider public interest in the advancement of learning, innovation and culture.

The ADA has a strong interest in ensuring balanced and appropriate international copyright policy, both due to its influence in Australia and to support the right of all humans to access knowledge.

In summary, we submit that the WIPO draft paper should:

- more explicitly acknowledge and seek feedback on the consequences of providing protection to machine generated works for access to knowledge, free flow of information, and the rights of human users and creators.

- more clearly define terms such as data and machine learning, to address the great variation of respondents' likely understanding of these terms;

- more clearly acknowledge and explore the role of the threshold of protection in determining the legality of data use, rather than just exceptions; and

- directly seek input on considerations of practicality in relation to licensing of data for uses such as machine learning.

We provide comments on select sections of the draft issues paper to demonstrate these points below.

Issue 6: Authorship and Ownership

We support the draft paper's acknowledgement that the provision of copyright protection for AI generated materials could have far reaching consequences for the copyright system, beyond the impact on specific works or businesses.

However, we are concerned that the introduction to Question 12 simplifies these potential consequences. It contrasts two scenarios for the copyright system - a system that favours the recognition of human dignity in creation and so does not protect AI creations versus a system that favours incentivising a large number of works and so does protect AI creations. In doing so, it considers only the impact that the recognition of AI works would have on the potential rights holders of the works in question, and does not consider its impact on users, other creators or access to knowledge in general.

A truly balanced and functional copyright system must consider not only its impact on incentives to create works, but also its impact on access to these works, on education, research and creativity that may be built on this access, and on the ability of downstream creators to create in the future. There is a global trend towards more overt recognition of access as being just as essential to functional IP systems as protection, as demonstrated by the rapid embrace of the Marrakesh Treaty and the increasing inclusion of language supportive of exceptions and limitations in bilateral and multilateral agreements.[1]

As Australia's own Productivity Commission has noted "when innovation is cumulative, IP rights can reduce the flow of benefits from new ideas and processes. . . Indeed, overly strong restrictions on diffusion can be so detrimental to innovation that it can 'undo' the benefits of the IP system in the first place."[2] They then go on to quote Nobel Prize in Economics winner, Joseph E. Stiglitz:

> a poorly designed intellectual property regime — one that creates excessively 'strong' intellectual property rights — can actually impede innovation. ... Knowledge is the most important input into the production of knowledge. Intellectual property restricts this input; indeed, it works by limiting access to knowledge. [3]

This is why, the Productivity Commission notes:

> The nature of costs and benefits — and how they change with 'greater' or 'lesser rights' — should also be considered when setting IP policy when evidence is mixed . . .The balance between creators and consumers of IP-intensive goods should also be kept in mind — an extension of rights without an extension of exceptions is difficult to justify except in extreme cases where IP rights are manifestly ineffective.[4]

The WIPO draft paper seems to acknowledge the potential negative effect of the extension of IP rights on the rights of consumers, educators, researchers and other creators under Issue 10, which discusses impacts on the

---

[1] See, for example, Article 18.66 of the Trans-Pacific Partnership Agreement, which states:
> Each Party shall endeavour to achieve an appropriate balance in its copyright and related rights system, among other things by means of limitations or exceptions that are consistent with Article 18.65 (Limitations and Exceptions), including those for the digital environment, giving due consideration to legitimate purposes such as, but not limited to: criticism; comment; news reporting; teaching, scholarship, research, and other similar purposes; and facilitating access to published works for persons who are blind, visually impaired or otherwise print disabled.

Available at https://dfat.gov.au/trade/agreements/in-force/cptpp/official-documents/Documents/18-intellectual-property.pdf
[2] *Intellectual Property Arrangements: Productivity Commission Inquiry Report No. 78* (Australian Government, 23 September 2016) p.4, available at https://www.pc.gov.au/inquiries/completed/intellectual-property/report/intellectual-property.pdf
[3] Stiglitz, J. 2008, 'Economic foundations of intellectual property rights', *Duke Law Journal*, vol. 57, pp. 1696-8. Available at https://scholarship.law.duke.edu/dlj/vol57/iss6/3/.
[4] *Intellectual Property Arrangements: Productivity Commission Inquiry Report No. 78* (Australian Government, 23 September 2016) p.73, available at https://www.pc.gov.au/inquiries/completed/intellectual-property/report/intellectual-property.pdf

"free flow of information". However, they are not mentioned in this earlier section, despite being equally or more relevant when discussing granting protection to a whole new category of materials.

The potential of AI enabled technologies to rapidly change the landscape of creation is unprecedented. In theory, it would be possible for an AI technology to generate innumerable new works almost instantaneously, with very little or no human input. Automatic protection of these works would place an unprecedented monopoly in the hands of a single player, or possibly a small number of players, most likely large corporations with the resources to develop and make use of such AI tools. Such a monopoly could easily be abused to restrict access to a huge amount of human knowledge for even the most beneficial of uses.

It is also easy to envisage a scenario in which the protection of AI generated works reduces the protection available to human authors, by essentially pre-empting them. For example, an AI technology could be tasked with determining all the possible combinations of the chromatic scale for a two minute musical piece. If each of these were to be protected as separate copyright works upon creation, we would see the position where subsequent works independently created by a human author could be shown to be a "copy" of an AI work, and so subject to licence. In a world in which ISPs and search engines receive millions of automatically generated takedown notices daily, with very high rates of inaccuracy,[5] and copyright trolls seek to extort settlements from private individuals with little knowledge of the law and no resources to seek legal assistance,[6] the protection of AI works could open the door to malicious attacks on publishers and creators alike by bad actors.

Such scenarios should be contemplated by WIPO in its consideration of the issue, and by responders in their consideration of this question. We therefore submit that Question 12 be amended to explicitly acknowledge the need to consider the impact recognition of autonomously generated works could have on access to knowledge and the rights of other creators.

Issue 7: Infringement and Exceptions

Question 13(i) - Should the use of the data subsisting in copyright works without authorization for machine learning constitute an infringement of copyright? If not, should an explicit exception be made under copyright law or other relevant laws for the use of such data to train AI applications?

We submit that this question does not adequately acknowledge the different forms of "data", the different ways that data can be used for machine learning, or the variations in the status of data under copyright law globally.

As the Berne Convention makes clear, it is a basic principle of copyright law that mere facts are not protected. The definition of "data subsisting in copyright works" therefore has a direct impact on the question of what is or should be protected - is this data a mere fact, or is it something more substantive? Different industries make different uses

---

[5] See this analysed in depth in a number of academic articles, including Urban, Jennifer M. and Karaganis, Joe and Schofield, Brianna, 'Notice and Takedown in Everyday Practice' (March 22, 2017). *UC Berkeley Public Law Research Paper No. 2755628.* Available at SSRN: https://ssrn.com/abstract=2755628 or http://dx.doi.org/10.2139/ssrn.2755628; Penney, J 'Privacy and Legal Automation: The DMCA as a Case Study' (22 STAN. TECH. L. REV. 412 (2019)) available at https://law.stanford.edu/wp-content/uploads/2019/09/Penney_20190923_Clean.pdf

[6] See settled case of *Youtube LLC v Christopher L Brady,* available at https://www.courtlistener.com/recap/gov.uscourts.ned.84712/gov.uscourts.ned.84712.1.0.pdf, summarised at Alexander, J. 'YouTube sues alleged copyright troll over extortion of multiple YouTubers' (The Verge, 19 August 2019) https://www.theverge.com/2019/8/19/20812144/youtube-copyright-strike-lawsuit-alleged-extortion-minecraft. The problem is also described in *McDermott v. Monday Monday* (S.D.N.Y. Feb. 22, 2018) available at https://scholar.google.com/scholar_case?case=10616859143516780806; at https://www.eff.org/issues/copyright-trolls; and https://www.illinoislegalaid.org/about/our-work/blog/beware-copyright-troll-letters-and-know-there-are-lawyers-who-can-help-you

of the term "data" and so form different understandings of where the boundaries of the term lie with some viewing it as akin to facts while others would include substantive materials such as tables and photographs. This is acknowledged in the introduction to the draft paper's Data section and in the questions raised under Issue 10, yet no definition is provided at any point throughout the paper.

Similarly, the concept of "use" for "machine learning" needs clarification to obtain an accurate answer from respondents. For example, merely reading a work in a non-consumptive manner does not constitute copyright infringement in many jurisdictions and respondents' answers are likely to be highly influenced by what they believe machine learning to involve.

Answering the question as currently phrased therefore becomes extremely difficult, as it covers a large range of scenarios on which respondents are likely to have differing opinions. Many respondents are likely to hold that the mere scraping of facts from a document does not constitute a protectable act under copyright law, whilst the reproduction of whole compiled databases would. They may also hold that digitally "reading" a work should not constitute an infringement, while using elements of it in a new work should.

We therefore propose that in this section WIPO:

■ provide additional guidance as to the definition of data it is using, and whether it applies to mere facts or more developed materials; and

■ provide a clearer definition of "use", including examples of the kinds of uses it contemplates the question covering.

On a related issue, it is equally important to note that exceptions are not the only method of limiting copyright protection. In many jurisdictions the legality of data use is government not through exceptions, but rather by consideration of the threshold of protection provided to materials and actions. As a consequence, asking respondents to comment on whether exceptions should be provided for data use, without also asking them about the settings that dictate when use of data will activate copyright, will provide an incomplete picture.

We therefore submit that Question 13 be broadened to also seek input on whether appropriate data access and use is best achieved by introducing exceptions, by clarifying the international standards around threshold of protection or by doing both.

Question 13(iii) - If the use of the data subsisting in copyright works without authorization for machine learning is considered to constitute an infringement of copyright, should an exception be made for at least certain acts for limited purposes, such as the use in non-commercial user-generated works or the use for research?

We submit that this question overlaps in a confusing manner with Question 13(i), which already asks whether an exception should be created to permit use of data in machine learning. We instead propose that this question be re-worded to follow on from 13(i) by focusing not on whether an exception should exist, but rather the scope of that exception eg by asking if an exception were made for machine learning whether it should be general or should be limited to certain purposes.

We also propose that "fairness" be added to the list of examples of potential limitations that could apply to an exception for machine learning. We note that existing TDM exceptions in countries such as the US, Canada and South Korea are based on a test of fairness rather than commerciality per se. We are also unaware of any exception which allows use of machine learning techniques only for "user-generated works".

Question 13 (v) - Would any policy intervention be necessary to facilitate licensing if the unauthorized use of data subsisting in copyright works for machine learning were to be considered an infringement of copyright?

Question 13 (vi) - How would the unauthorized use of data subsisting in copyright works for machine learning be detected and enforced, in particular when a large number of copyright works are created by AI?

We welcome both these questions, as they implicitly acknowledge the impracticality of requiring direct licensing of works for the purpose of machine learning. However, we submit that this impracticality should be acknowledged more explicitly in the questions set out under Issue 7.

In the realm of machine learning the large numbers of works needed to be used makes it effectively impossible to identify and contact all copyright owners. Even if copyright owners are able to be identified, the contribution of each individual work will generally be so small as to make payouts insignificant, and far outweighed by the transactional costs of administering the licences. Such practical problems also affect compulsory licensing schemes. Indeed, we have seen their effect in the orphan works licensing schemes introduced in the UK and Canada, with both systems proving expensive to set up and run, difficult to navigate, and bureaucratic, whilst providing very little compensation to rights holders. In such circumstances the only beneficiary is the collective licensing entity, and the scheme itself becomes a tax on innovation and learning rather than an incentive to create.

We therefore urge WIPO to add a separate question about the practicality of using licensing to permit machine learning, and the effect on the free flow of information, innovation and technological development if such an approach were to be relied upon. This question could then be followed up with questions regarding compulsory licensing schemes and the practicality of enforcement.