UK Consultation on Artificial Intelligence and Intellectual Property Policy

Submission by Creative Commons

[redacted][1]

November 30, 2020

Introduction

This document is submitted to the UK Intellectual Property Office (UKIPO) consultation on copyright and related rights. It summarizes Creative Commons' provisional position on matters relating to artificial intelligence (AI) and copyright law. CC is dedicated to facilitating greater openness for the common good. In general, we believe that the use of publicly available data on the Internet leads to greater innovation, collaboration, and creativity.

Generally, in view of the uncertainty surrounding the concept of AI, Creative Commons:

- supports broad and unfettered access and use of copyright works as AI input, in service of the public interest, in particular through mandatory exceptions and limitations, and bearing in mind any limitation arising out of ethical or privacy concerns;
- opposes copyright (or related right or any new *sui generis* right) protection of AI-generated outputs, since relying on copyright to govern AI output is contradictory to copyright's primordial function of offering an enabling environment for human creativity to flourish.

Creative Commons looks forward to considering the opinions and ideas contributed by others in connection with the consultation as it refines its position in connection with the UK IPO's work.

1. **Lack of clarity around AI - Copyright intervention must be done with caution**

[1] This paper has been prepared by [redacted] It has benefited from invaluable input from Creative Commons staff, including [redacted]. Creative Commons is grateful to members of the Creative Commons community and partners who contributed and provided insightful comments on this submission; however, this submission does not necessarily represent their opinions or those of the institutions with which they are associated, namely: [redacted] (Communia/Intellectual Property Institute), [redacted] (Wikimedia Deutschland), [redacted] (Communia), [redacted] (Communia), [redacted] (Communia/Centrum Cyfrowe) and [redacted]

- Clarity on basic definitions in the AI space is a prerequisite to competent regulation in the copyright arena. AI needs to be properly understood before any copyright implications can be addressed.
- Any policy or legal intervention in the field of copyright should be based on strong and reliable evidence and conceptual certainty, especially given the fast-paced evolution of AI technology.
- It is presently not clear what AI exactly is and what it is capable of producing.[2] As things stand, the term "AI" is not defined precisely enough to be used in the copyright arena. At the very least, any document dealing with AI should provide or refer to a clear and precise definition of the term.
- Given this lack of clarity on what AI really means, intervention in copyright law and policy must be carried out with caution.

2. Pressing policy issues outside the copyright arena need addressing

- Many issues raised by the development of AI are to be addressed under the lens of ethics, cultural rights and interests, fundamental human rights (including the principles of equality and non-discrimination), personality rights, privacy rights, and data protection.
- These adjacent issues should be addressed and debated in their respective policy arena, and must inform but not encroach on the framework of copyright.
- Consideration of the copyright implications can be undertaken while bearing in mind that these issues need to be satisfactorily addressed and resolved in their own policy sphere.
- Noting that those other issues are key to a coordinated and inclusive policy approach on AI and that they have a direct impact on any copyright discussions on AI, the present submission is nevertheless limited to the substantive questions that AI raises in the copyright arena.

3. Copyright is not the right framework to regulate AI outputs

- Copyright (including related rights and *sui generis* rights) is not a universal tool meant to address any and all problems that society might encounter.
- Using copyright to govern AI outputs is contradictory to copyright's primordial function of offering an enabling environment for human creativity to flourish. It should not be

[2] "AI" itself is an evolving concept. AI is an umbrella term that encompasses, for the most part at the current time, different types of machine learning algorithms. There is a lot of confusion around related and different concepts such as machine learning, natural language processing, predictive models, neural networks as well as algorithms. AI is generally understood as "something that can be done by a computer that until then, could only be exclusively done with human intelligence." However, that new capability is only "AI" until it becomes normalized as simply "software." What is AI today may not be so tomorrow. As technology advances over time, what is considered "AI" as opposed to "normal software" may continually evolve. That means that whatever copyright framework is put into place, if any, it has to remain flexible enough and technology-neutral to account for and adapt to the moving-target nature of AI. There is also danger in categorizing all manner of algorithms as "AI" and in adopting rules or measures where these categories are arbitrarily determined.

assumed that a system developed to protect human creative expressions can simply be applied to non-human output.
- Copyright has yet to adapt to the digital environment, which emerged several decades ago; it is unwise to attempt to force the application of an already outdated system to the nascent and uncharted field of AI technology.
- To the extent that copyright might be deemed relevant to regulate AI output, this policy brief provides guidance on the basic questions to consider.

Part I. No copyright for AI outputs

- Although AI might arguably generate output akin to a copyright work, this output does not necessarily meet the copyright protection requirements, such as authorship and originality.
- AI-generated outputs are not "works" as this term is defined in copyright law.
- Granting copyright to AI output would raise important concerns around intellectual property rights overlap leading to overprotection. Other exclusive rights (including copyright and patent rights in AI software) already provide sufficient protection in the AI space. Overprotection can have negative impacts on creativity, innovation and the provision of public goods.
- Introducing new rights complicates an already complex field of copyright, related rights and *sui generis* rights, and risks entangling the exercise of existing and any future exceptions and limitations.
- AI-generated content should not benefit from any copyright (economic or moral rights), related rights or *sui generis* protection.

1. Generating content through AI is not an act of authorship

a. Persons who develop AI algorithms are not the authors of AI-generated outputs.

- The persons (physical or legal) who program the AI should not be entitled to copyright in the output generated by the program they created.
- AI-generated output is made of a mechanical conglomeration of the inputs. Absent direct human intervention or direction, this process lacks the authorial contribution necessary to warrant copyright protection.
- Content generated without human creative input should be in the public domain.[3]

b. AI is not capable of producing content "autonomously"

[3] On this point, see Gervais, Daniel J., "The Machine As Author" (March 25, 2019). Iowa Law Review, Vol. 105, 2019; Vanderbilt Law Research Paper No. 19-35.

- There is uncertainty about whether and to what extent AI is capable of producing content "autonomously" without any direct, material human involvement.
- It is not established that AI can acquire and develop autonomous and cognitive features through experience learning that are sufficient to generate creative output. A certain level of human input is always required.
- The mechanistic nature of AI algorithms bars any possibility of autonomous creation. The output of any current AI algorithm is essentially a mathematical function of its inputs, be they image or video files or textual data. This is not a valid basis for claims of autonomous authorship.

c. AI systems are not authors

- The notion of the human creator, the human creative spirit and human creativity are bedrock principles of the copyright system.
- Direct human, authorial involvement is a precondition to determining whether a work is worthy of copyright protection or whether copyright can be claimed.
- Bestowing copyright upon AI-generated outputs to the copyright or patent holder of the AI software, algorithm or program, or to any of its users or to contributors of data, is not advisable absent evidence of direct human creativity[4] and in light of concerns about over-protection.
- One would be mistaken to presume an equivalence between human-authored works and AI-generated outputs. AI and the copyright principle of authorship are antithetical concepts; stated otherwise, "the very idea of AI authorship is oxymoronic."[5]
- The generation of output through an AI process without human involvement is not an act of authorship.

d. AI-generated output is not likely to be considered "original"

- It is not clear how to judge the originality of a work essentially composed of random snippets of thousands or millions of input works.
- AI-generated outputs should not as a default be considered original works. Originality is a reflection of the intellectual, creative choices made by a human author.
- Output generated solely as a result of a mathematical function lacks originality.

---

[4] The users of programs who contribute to the AI outputs may be entitled to copyright in the output generated by the program they use if they contribute with substantial and sufficient creative input. On a similar point, see Samuelson, Pamela, "Allocating Ownership Rights in Computer-Generated Works" *U. pitt. L. rev.* 47 (1985): 1185. The contributors of data that are used in the process of machine learning which ultimately lead to AI-generated output should not be entitled to copyright in the output generated in such processes absent any direct, creative, authorial involvement.

[5] Craig, Carys J. and Kerr, Ian R., The Death of the AI Author (March 25, 2019). Osgoode Legal Studies Research Paper, at 5.

2. AI cannot have human rights

- The Universal Declaration on Human Rights (UDHR) protects the moral and material interest of authors resulting from scientific, literary or artistic production and as such recognizes authors' rights - rights deriving from human authorships as human rights.
- AI systems cannot have human rights.

3. AI cannot be held liable for its actions

- Authors are granted rights in their creations; in return, they are liable for their actions.
- AI itself cannot be held liable for copyright infringement nor for any other breaches of rights caused by its generated output, due to a current lack of regulation of AI liability. Since "copyright (as in a right in one's work) and responsibility for that work historically have gone hand in hand,"[6] AI-generated output should not be afforded copyright protection.

4. AI development does not require copyright in AI-generated output as an incentive

- Producers of AI do not require copyright protection in AI-generated outputs as an incentive to develop AI.
- Market opportunities are already rife, in the absence of any copyright protection for AI-generated outputs.
- The efforts invested in the creation or invention of the AI program are to be incentivized are rewarded through means other than copyright protection over the AI-generated output. Incentives and rewards in recognition of the investment made and the innovation brought about by the organizations and individuals involved in the development of AI can be found in other areas, including copyright in the code of the AI software itself[7], as well as patents, trade secret laws, and laws protecting against unfair competition.
- Copyright's utilitarian doctrine and incentives theory cannot support a claim that AI be afforded rights for any generated output because AI fails to meet the role of the author and its contribution to human-led social progress.[8] Granting AI-outputs the status of copyright work goes against the social purpose for which copyright was created.

5. Very limited protection for AI outputs as a fall-back option

---

[6] Gervais, Daniel J., "The Machine As Author" (March 25, 2019). Iowa Law Review, Vol. 105, 2019; Vanderbilt Law Research Paper No. 19-35, at 60.
[7] There is evidence that open software and sharing cultures are thriving in the AI space.
[8] On this point, see generally Hilty, Reto and Hoffmann, Jörg and Scheuerer, Stefan, "Intellectual Property Justification for Artificial Intelligence" (February 11, 2020). Draft chapter. Forthcoming in: J.-A. Lee, K.-C. Liu, R. M. Hilty (eds.), Artificial Intelligence & Intellectual Property, Oxford, Oxford University Press, 2020, Forthcoming; Max Planck Institute for Innovation & Competition Research Paper No. 20-02.

- Assuming the undesirable outcome that new copyright, related rights or *sui generis* rights would be established to regulate AI and protect AI-generated content despite our strong opposition, this should be done conservatively and with restraint.
- Considerations should include: a high bar for the creation of such new rights, a short term of protection, and robust exceptions and limitations to uphold users' rights, safeguard the public interest, and ensure a vibrant public domain.

6. No new *sui generis* rights for AI-generated content

- There should be no new *sui generis* rights established for AI-generated content.
- The incentives that already exist within the copyright (for AI software) patent, trade secrets, and unfair competition systems are sufficient to encourage the development of AI applications.

Part II. Use of copyright material as AI input/training

- The use of copyright works as input or to train AI applications should not necessarily be considered copyright infringement as a default. It should be generally allowed under clear and open exceptions and limitations where such use upholds the public interest.
- Unfettered access and use of data to improve and build upon AI encourages innovation and development of AI in support of public-interest activities. It helps reduce bias, enhance inclusion, and promote important activities such as education and research.
- That said, other concerns must be taken into account when using material to train AI. For instance, one must consider the tension between the value of open data and legitimate concerns about privacy as well as ethical and responsible use of copyright material (especially openly licensed content) for algorithmic training.

1. Openly licenced content

- The open access movement demonstrates the obvious advantages of freely and openly accessible resources to spur innovation, especially in times of crises.[9] AI innovation is bound to be stimulated by openly accessible materials.
- In the context of openly CC-licensed content used to train AI applications, from a copyright perspective, CC has determined that no special or explicit permission regarding new technologies is required from the licensor. Creative Commons' FAQs[10]

[9] Victoria Heath and Brigitte Vézina, "Now Is the Time for Open Access Policies—Here's Why," Creative Commons Blog, March 19, 2020, https://creativecommons.org/2020/03/19/now-is-the-time-for-open-access-policies-heres-why/.
[10] Creative Commons, FAQs, "Artificial Intelligence and CC Licenses," https://creativecommons.org/faq/#artificial-intelligence-and-cc-licenses

clarify how the CC licenses work in the context of openly licensed content that is used to train AI tools.

- As a framework, CC licenses do not restrict reuse to any particular types of reuse, so long as the license terms are respected, as the case may be. Also, CC licenses do not override limitations and exceptions.

2. Exceptions and limitations

- Speaking of copyright-protected data generally — whether released under an open license or not — assuming access to copyright works is legitimate at the point of input, use of data to train AI should be considered non-infringing by default. This use is non-expressive and does not compete with the original works in any market. For example, text-and-data mining for research or education purposes should be allowed under an exception to copyright.
- Limitations and exceptions for cross-border collaboration on AI can foster creativity, innovation and the public interest, such as for education and research purposes, and contribute to international development.[11]
- Licensing, including collective licensing, is not an appropriate alternative to a system of exceptions and limitations upholding the public interest to enable the use of copyright works as AI input.

a. Text and data mining and the right to research

- Text and data mining (TDM) activities are pivotal in supporting research and innovation and in the training of AI systems.
- TDM activities are non-consumptive and non-expressive uses of work. TDM does not compete with original markets for works, and may indeed enhance them by increasing demand for a wider range of works.
- TDM should not be made subject to additional authorisations or payments once access is legitimate. Generally, TDM activities should not be considered copyright infringement and should not be restricted by copyright.
- TDM should be allowed and supported pursuant to exceptions and limitations, in particular to enable a proper exercise of the right to research and AI training-related activities.[12]

b. Bias

- We encourage the use of larger and more diverse sets of data in order to avoid bias in outputs.

---

[11] On this point, see Flynn, Sean and Geiger, Christophe and Quintais, João and Margoni, Thomas and Sag, Matthew and Guibault, L. and Carroll, Michael W., “Implementing User Rights for Research in the Field of Artificial Intelligence: A Call for International Action” (April 20, 2020). European Intellectual Property Review 2020, Issue 7.
[12] *Ibid.*

- Broad and open exceptions and limitations should apply to support the most extensive possible use of copyrighted works as AI input in order to encourage the elimination and minimization of bias.[13] Placing barriers around copyright material that can be freely mined risks increasing the likelihood of AI bias, unfairness and exclusion.
- One way to reduce bias, unfairness and exclusion in AI systems, other than ensuring that the algorithm itself is not biased, is to ensure that the maximum volume and widest diversity of content is available for training purposes, requiring both minimising unnecessary barriers to TDM and facilitating uses across borders.
- A careful balance must be struck between a push to reduce bias, unfairness and exclusion in AI on the one hand and privacy rights and ethical and human rights considerations on the other.

3. Digital rights management/Technological protection measures

- There should be no digital rights management (DRM) or technological protection measures (TPM) to restrict or prevent (otherwise legal) access to the data. There should be ethical requirements for transparency in the modalities of use of data, however this should be established outside the boundaries of the copyright system.

4. Concerns over data used for objectionable purposes

- When data is used for objectionable or problematic purposes (facial recognition for instance), privacy and surveillance concerns must be taken seriously.
- Nevertheless, copyright is not the most appropriate area of the law to protect individual privacy, to address research ethics in AI development or to regulate the use of surveillance tools employed online.
- That said, copyright discussions cannot take place in a policy vacuum. To arrive at a good, comprehensive solution we need to fully capture the implications and concerns raised in different policy arenas.
- The conversation on AI policy must be held in a coordinated and inclusive manner and through the lens of ethics, responsibility and sustainability, cultural rights, human rights, personality rights, privacy rights, and data protection. These other issues deserve equal attention and should not be marginalized.
- At CC we intend to reflect upon these adjacent issues. As with any fundamental ideal, the “openness” of data is not an absolute end in itself and must be balanced with considerations for privacy and ethics. Whether this balance is found *intrinsically* within the copyright system or *extrinsically* in other areas of the law, and whether the CC licenses and tools have any role to play, is something we look forward to probing.

[End of document]

[13] On this point, see generally, Levendowski, Amanda, “How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem” (July 24, 2017). 93 Wash. L. Rev. 579 (2018).