

Electronic submission via ai2ip@wipo.int

February 14, 2020

Subject: WIPO Conversation on Intellectual Property (IP) and Artificial Intelligence (AI): WIPO/IP/AI/2/GE/20/1

The Association of American Publishers (AAP) appreciates this opportunity to provide its views in this WIPO request for comments.

AAP is the national trade association that represents the leading book, journal, and education publishers in the United States on matters of law and public policy, advocating for outcomes that incentivize the publication of creative expression, professional content, and learning solutions. The U.S. publishing industry, indeed, the global publishing industry, supports an extensive network of multinational businesses and thousands of jobs across the globe.

AAP has a particular mandate and expertise in copyright law, seeking to promote an effective and enforceable framework that enables publishers and their technology partners to create and disseminate original works of authorship through ever-evolving business models to the benefit of their customers and the worldwide public. For this reason, our comments below focus on the potential impact of artificial intelligence (AI) capabilities on the established principles and application of copyright law and policy.

We live in a technologically abundant world, with new advances in technology and their application to myriad human tasks occurring rapidly. Huge growing stores of digitized data, combined with constantly increasing computational power and applications to analyze, organize and otherwise utilize data, are making the use of predictive AI a reality that promises to bring significant benefits to humankind in agriculture, financial services, transportation, healthcare, manufacturing, education, national defense, and countless other areas of human life.

Yet, while we hurtle forward in eager anticipation to embrace its presumed rewards of innovation and progress, the policies that should govern the development and application of AI in many, if not most, of these areas remain unclear. This is no less true for the impact that AI may have on the application and significance of copyright for creators, owners, distributors, and users of works of original expression. There are multiple questions to be addressed, with numerous potentially huge consequences. How will the exclusive rights of copyright owners be recognized and enforced when their copyrighted works are used to train AI systems? What ethical parameters should direct AI development, including with respect to authorship of works of original expression? How will the use of AI technologies inform and impact the private and public lives of creators, owners, distributors, and users of copyrighted works? WIPO's effort to crystalize the questions that should be asked and to

frame the issues that should be discussed with respect to intellectual property in general and the specific domain of copyright is both timely and important.

The WIPO Conversation document notes that “[a]rtificial intelligence” (AI) has emerged as a general-purpose technology;” yet, the document does not attempt to articulate a more specific definition of artificial intelligence, despite the absence of a sufficiently settled definition of the term to meaningfully guide WIPO’s inquiry. While there is no single agreed definition of AI, various extant usages of the term appear to coalesce around the notion of AI as data-driven technologies which cannot be fully severed in their possibilities and consequences from the human actors who design them or facilitate their ingestion of particular datasets for instruction on accomplishing particular tasks or achieving particular results.

While WIPO’s interest in AI and its implications on the intellectual property framework is commendable and to be expected, WIPO should – at least for the present -- limit its interest to gathering information on the various inquiries and discussions occurring around the topic, and stay away from any premature determinations at the international level on whether a new regulatory regime is necessary to address the use of data embodied in copyrighted materials to train AI algorithms or processes.

It should also be noted that, as WIPO is the primary international agency charged with promoting a strong intellectual property (IP) protection framework, a first principle underlying these discussions is that the exclusive rights of creators and owners of copyrighted works should be respected and protected. It remains WIPO’s role to promote IP protection, particularly as the race for supremacy in developing AI technology may drive governments to ignore the importance of IP rights. AI development and training is done by feeding huge troves of data to an AI algorithm, which may in many instances implicate valuable and proprietary copyrighted works. Whether the use of data subsisting in copyrighted works¹ to train AI should be permissioned and compensated must necessarily be evaluated under existing copyright law frameworks, which must comply with current international norms. It is also important to note that any legally-actionable definitions of “data” – crucial to AI training - should be appropriately narrow so as to not inadvertently, i.e., without intention to apply key principles and rights of copyright, include copyrighted works, such as books, journal articles, and other creative works developed for and created by private sector rights holders.

For the purpose of informing WIPO’s conversations around AI, we address below the issues posed in the section on Copyright and Related Rights, which are most pertinent to publishing interests.

Issue 6: Authorship and Ownership

- (i) Should copyright be attributed to original literary and artistic works that are autonomously generated by AI or should a human creator be required?**

¹ For AAP, the phrase “data subsisting in copyrighted works” and “data embodied in copyrighted works” carries the same meaning as the phrase “copyrighted works treated as data.”

The Berne Convention provides copyright protection for the works of creators, including authors, musicians, and painters. Under the Berne Convention, and the copyright laws of many jurisdictions around the world, including the United States, the “creator” or “author” is envisioned as a natural person. Thus, copyright cannot be attributed to an original literary and artistic work that is autonomously generated by AI, entirely independent of a human creator. Whether independent AI-creation may qualify as “subject matter other than works,” as is formulated in some jurisdictions, is a question that should perhaps be addressed as this inquiry progresses.

- (ii) In the event copyright can be attributed to AI-generated works, in whom should the copyright vest? Should consideration be given to according a legal personality to an AI application where it creates original works autonomously, so that the copyright would vest in the personality and the personality could be governed and sold in a manner similar to a corporation?**

As noted above, it is AAP’s view that copyright cannot be attributed to literary or artistic works autonomously created by AI, as human involvement in the creative process remains a requirement under international copyright law. While advances are inevitable, AI technology is not yet at a stage where engaging in the legal fiction of granting legal personality to an AI application would appear to be necessary or appropriate. In any case, while the line may be long, that line - between the human originator of the AI algorithm’s programming and the AI’s creative output - can still be drawn. As such, the creative element can be ascribed to the human originator to whom copyright ownership, and therefore control over the attendant rights, may be accorded.

We should also consider the purpose of the grant of copyright, which is both to promote the economic rights of creators and, by allowing them to receive such benefits from their intellectual creations, to incentivize them to create and disseminate more works to inform, educate, entertain, and otherwise provide broader societal benefits to individuals, communities and the general public. As these same policy considerations would not seem to readily apply to an AI algorithm or process, the rationale for the grant of copyright to an autonomous AI creation does not apply, making the grant of legal personality to an AI process both unnecessary and inappropriate.

- (iii) Should a separate *sui generis* system of protection (for example, one offering a reduced term of protection and other limitations, or one treating AI-generated works as performances) be envisaged for original literary and artistic works autonomously generated by AI?**

As the existing (international) copyright law framework does not neatly address the question of copyrightability of autonomously generated AI-creations, the development of a *sui generis* system of protection for independent AI-creations warrants exploration. Yet, it should be recognized that the perceived need for such a framework is solely for the purpose of addressing the question of the copyrightability of independent AI-creation. If the view is taken that the rationale for according copyright protection to human creation does not apply with respect to AI algorithms or processes creating works, the question arises as to why the creation of a *sui generis* framework is even necessary. Indeed, the more important questions appear to be around the issues of ownership of the independently created AI-

work, and how its uses are to be managed – for which the law of contract or licensing may already provide appropriate and workable answers.

Issue 7: Infringement and Exceptions

- (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?**

The above question cannot be answered in the abstract. Whether and how a particular use of data subsisting in copyrighted works without authorization for machine learning constitutes an infringement of copyright is best addressed through the lens of the relevant national copyright law, provided that such national law is compliant with international copyright law standards.

Though the question cannot be answered in the abstract, it nonetheless remains AAP's view that wholesale, un-permissioned reproduction of copyrighted works in which data subsists, even for the purpose of machine learning, is likely to be infringing. Where data embodied in copyrighted works is to be used for machine learning purposes, the scope and terms of such use can best be set out in a licensing agreement between the parties. Licensing remains the most flexible tool through which AI training can be promoted, while also recognizing and protecting the copyrights of rights holders.

It is worth noting that, notwithstanding the lack of a specific exception for AI training in many jurisdictions, commercial and non-commercial entities are already engaging in AI training activities. Usage of data embodied in copyrighted works for machine learning purposes is already ably facilitated through licensing agreements or contracts between the data user and the owner(s) of the copyrighted works in which data may subsist. The fact that these arrangements already exist show that many current national copyright law frameworks are not a hindrance to AI development and enrichment, thereby negating any perceived need for creating new exceptions and limitations purportedly to satisfy the purpose of AI training.

- (ii) If the use of the data subsisting in copyright works without authorization for machine learning is considered to constitute an infringement of copyright, what would be the impact on the development of AI and on the free flow of data to improve innovation in AI?**

Respecting the exclusive rights of the copyright owner in the copyrighted works, which may embody data, is not a restriction or impediment to the free flow of data. Promoting the free flow of data means the removal of barriers such as data localization requirements, facilitating data accessibility, or ensuring the security of data flows. Ensuring that rights holders are duly compensated for particular uses to which an AI-research entity may wish to put curated or proprietary data, in which the rights holder has invested, is an appropriate recognition and exercise of legitimate economic rights. The recognition and exercise of these rights facilitates, rather than restricts, data flows. Likewise, as noted

previously, whether a particular use is infringing can only be adjudged according to the particular facts of a case. AI development and training will not be impeded by a finding of infringement, as the infringing party can cure the infraction by securing the appropriate permissions or licenses to use the copyrighted works in which data may be embodied.

- (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 use for research?**

We refer to our response under Issue 7 (i) on this matter. In short, we do not find that a need for such an exception exists, and strongly caution against the creation of new exceptions to purportedly facilitate the use of data subsisting in copyrighted works for machine learning or AI training purposes. Licensing solutions remain the best tool for facilitating AI development while also protecting the rights of creators, publishers, and other copyright owners and licensees. Licensing arrangements will provide the desired flexibility, while affording rights holders and users of data greater stability and certainty with respect to their rights and obligations.

- (iv) If the use of data subsisting in copyright works without authorization for machine learning is considered to constitute an infringement of copyright, how would existing exceptions for text and data mining interact with such infringement?**

Where an exception for text-and-data mining (TDM) is in place, whether a particular use qualifies under this exception must obviously be evaluated under the intended meaning of the language of the exception within the relevant jurisdiction. Per such explicit exception, the otherwise infringing act of reproducing entire bodies of work for certain TDM activities is deemed non-infringing if the requirements for the exception's application are met – such as the nature of the person or entity seeking to engage in TDM activities, the purpose of the TDM exercise, and the uses to be made of the output of such TDM activity.

The EU Directive on copyright in the Digital Single Market (DSM) created a bifurcated TDM exception, in which it differentiated between TDM for commercial and for non-commercial purposes (specifically, for scientific research). Where the user (i.e., the text-and-data “miner”) is a commercial or for-profit entity, the DSM requires that commercial user to obtain a license “for reproductions or extractions” of “works or other subject matter to which they have lawful access,” where such use of the works has been reserved by the rights holder.

- (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?**

No, such an intervention is unnecessary. Again, we refer to our responses above to Issue 7 (i) and (iii). Licensing is already enabled by the existing copyright law frameworks of numerous jurisdictions. The licensing framework, of course, is premised on a strong copyright protection regime, from which rights holders derive their ability to protect and

exploit their rights in the works (economic and otherwise), that may be treated as data desirable for use in AI training.

(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?

The answer may lie in technology. Rights holders may have to embed rights management information (RMI) technology in works or collections of works that define or specify what can and cannot be done with the works. For instance, in the EU DSM, Article 4 (3) provides that rights holders may reserve their rights in an “appropriate manner, such as machine-readable means in the case of content made publicly available online.” National authorities may also consider requiring entities engaged in AI research to document the source of the data used, which may also include information regarding the permissions obtained and other licensing parameters.

Issue 8: Deep Fakes

(i) Since deep fakes are created on the basis of data that may be the subject of copyright, to whom should the copyright in the deep fake belong? Should there be a system of equitable remuneration for persons whose likenesses and “performances” are used in a deep fake?

Issues arising from so-called “deep fakes” are not easily evaluated nor solely to be evaluated within the copyright law framework. “Deep fakes refer to manipulated videos, or other digital representations produced by sophisticated artificial intelligence, that yield fabricated images and sounds that appear to be real.”² The more profound issues of personal identity, the right to privacy, the right of publicity, and the ability to control the use of one’s image for any purpose appear more appropriately to be human rights issues, rather than purely or even primarily copyright issues. Indeed, the question should perhaps be whether copyright should even be accorded to deep fake imagery, rather than to whom copyright in a deep fake should belong. If the deep fake imagery depicts a human subject in a manner or light wholly inconsistent with the subject’s life, life’s work, or status, it seems incongruent that this deep fake should be rewarded with copyright protection.

On the other hand, it may be helpful to consider whether there are instances where deep fake imagery may be deserving of copyright protection. In such a case, the copyright might properly belong to the human actor(s) from whom the design and function of the AI program that creates the imagery originates. For example, an audiovisual producer may develop an AI program in-house to re-create the image of a deceased actor for its use in a new film. The copyright in the resulting deep fake may be accorded to the audiovisual producer. Yet, it might also be the case that the deep fake may be produced utilizing a commercially available AI algorithm, where the human actor uses the AI algorithm to accomplish his creative vision in much the same way as a photographer uses a camera to bring forth his perspective. Copyright ownership, in this latter case, could be accorded to the human actor employing the AI algorithm as a tool.

² See <https://www.cnbc.com/2019/10/14/what-is-deepfake-and-how-it-might-be-dangerous.html>

Issue 9: General Policy Issues

- (i) **Are there seen or unforeseen consequences of copyright on bias in AI applications? Or is there a hierarchy of social policies that needs to be envisaged that would promote the preservation of the copyright system and the dignity of human creation over the encouragement of innovation in AI, or vice versa?**

It is important that the rationale for according copyright protection be maintained – i.e., that the human element in creating any work of original expression is essential to the grant of copyright. An AI algorithm, while it may eventually achieve a sophistication that enables it to independently create, is ultimately likely only fulfilling its programming imperative – having been designed and trained by human actors to achieve a desired task. The AI algorithm will have been trained on the data (e.g., images, writings, sound recordings, and other creative works) that its programmers choose to feed it. This process of “training” would seem to indicate that the AI algorithm is merely emulating the nature of the multitude of data points on which it has been trained, rather than independently creating based on its interactions with and experiences of the world around it – which is what drives the human creative process. Viewed in this light, the argument is all the more urgent and compelling for why the dignity of human creation, and the copyright framework that fuels such creation, should not be sacrificed on the altar of AI development. It is the human spirit and its unique spark of ingenuity that drives the creation of original expression, and this spark cannot yet be emulated by any machine or algorithm. Human creative genius should not be cheapened by rhetoric that would degrade copyright, such that creative works are to be treated merely as fodder for AI development and training.

Issue 10: Further Rights in Relation to Data

- (i) **Should IP policy consider the creation of new rights in relation to data or are current IP rights, unfair competition laws and similar protection regimes, contractual arrangements and technological measures sufficient to protect data?**

There does not seem to be a need to create new rights in relation to data – at least under the copyright framework. Indeed, the creation of a new IP right for data would seem contrary to the fundamental principle that facts and ideas do not qualify for copyright protection.

If the establishment of new proprietary rights in data is to be seriously contemplated, consideration should first be given as to whether existing legal frameworks *other than copyright* already serve the purpose of providing necessary and appropriate protection(s) for data or warrant development for that purpose. There is no reason to risk distortion of established principles of copyright to accommodate data protection concerns, no matter how legitimate those concerns may be. It would be essential, therefore, as a threshold matter, to identify both the type(s) of data deserving of protection and the type(s) of protection they should be afforded.

For example, it would be beneficial for individual dignity to enable all persons (or the data subject) to assert ownership and other legal rights over their personal data, such as

the ability to control who may access this data, and how or for what purposes personal data may be used. In Europe, for example, the General Data Protection Regime (GDPR) provides individuals with tools to allow them to control their personal data, such as those necessary to exercise their privacy rights or the right to control how their personal data is processed. Given that technology companies already harvest significant quantities of personal data, often without the explicit knowledge or consent of individual consumers, frameworks that enable individuals to assert control over their personal data are of critical importance. This, however, is not a copyright issue.

For non-personal data, such as proprietary data that may be the result of confidential research or the effort and investment by an organization to obtain, verify, or curate (factual) information, contract law may be the most robust means through which database owners may assert, exercise, and protect their rights to control access to and uses of such data. As the above question implies, contractual arrangements between parties can adequately address issues such as scope of use of proprietary data (e.g., restrictions as to extraction and re-use), how and with what persons or entities proprietary data may be shared, and measures that the user may be required to put in place to ensure security of the proprietary data. Certainly, technological protection measures may be useful to ensure such contractual terms are adequately enforced; moreover, such measures may be even more important when contractual arrangements would be problematic due to a lack of legal privity between database owners and those seeking to use the database and its contents.

Nonetheless, while underlying data may not be the subject of copyright, compilations of data or datasets may be copyrightable – though how such data and datasets are protected will necessarily vary by jurisdiction.

For instance, in the U.S, there is no *sui generis* protection for databases or data sets akin to that available under EU Directive 96/9/EC on the legal protection of databases. U.S. law does, however, protect data collections, or “compilations,” where authorship inheres if the compiler has “selected, coordinated, or arranged” information “in such a way that the resulting work as a whole constitutes an original work of authorship.”

To the extent that copyrighted works may be treated as data, it is critical that, where such “data” exists as a compilation curated by a rights holder, ingestion of entire databases to train AI must be permissioned and/or compensated, or otherwise compliant with copyright law. In jurisdictions where a database protection right already exists, the interplay of the database owner’s rights with the privileges granted to a potential user will be determined by the applicable laws, which may be copyright law, in the relevant jurisdiction.

Finally, as implied by previous statements throughout these comments, given the importance of data to AI training, AAP wishes to emphasize how critical it is for any legally-actionable definitions of “data” to be appropriately narrow so as to not inadvertently – i.e., without intention to apply key principles and rights of copyright -- include copyrighted works, such as books, journal articles, and other creative works developed for and created by private sector rights holders.

- (ii) If new IP rights were to be considered for data, what types of data would be the subject of protection?**

See response under Issue 10 (i).

- (iii) If new IP rights were to be considered for data, what would be the policy reasons for considering the creation of such rights?**

See response under Issue 10 (i).

- (iv) If new IP rights were to be considered for data, what IP rights would be appropriate, exclusive rights or rights of remuneration or both?**

As noted in the response to Issue 10 (i), for a variety of reasons, the copyright framework has not been generally embraced as the appropriate regime under which data can or should be protected.

- (v) Would any new rights be based on the inherent qualities of data (such as its commercial value) or on protection against certain forms of competition or activity in relation to certain classes of data that are deemed to be inappropriate or unfair, or on both?**

See response under Issue 10 (i)

- (vi) How would any such rights affect the free flow of data that may be necessary for the improvement of AI, science, technology or business applications of AI?**

See response under Issue 7 (ii).

- (vii) How would any new IP rights affect or interact with other policy frameworks in relation to data, such as privacy or security?**

See response to Issue 10 (i).

- (viii) How would any new IP rights be effectively enforced?**

See response to Issue 10 (i).

Conclusion

We appreciate WIPO's efforts to facilitate discussion of the implications of AI on intellectual property, particularly copyright policy. Data is essential to the development of AI technologies, but it will in many instances be embodied in the copyright protected works of authors, publishers, and other copyright owners. Policies directed at facilitating AI development must be such that its pursuit does not unreasonably impinge on nor detract from the rights of creators and rights holders in whose works may be embodied the data needed to train AI. The overall ecosystem must remain balanced and rational.

While this further inquiry will be useful to informing WIPO's distilling of the issues surrounding AI development and the use of copyrighted works to facilitate such development, further discussions of this complex issue will remain necessary. We thank WIPO for the opportunity to respond to this request for comments and look forward to participating in further conversations on this important issue.

Sincerely,

A handwritten signature in blue ink, appearing to read "M. Luisa Simpson". The signature is fluid and cursive, with a prominent initial "M" and a long, sweeping underline.

M. Luisa Simpson
Senior Vice President, Global Policy