AAP | ASSOCIATION OF AMERICAN PUBLISHERS

Electronic submission via AIPartnership@uspto.gov

January 10, 2020

The Hon. Andrei Iancu
Under Secretary of Commerce for IP
Director, U.S. Patent & Trademark Office
P.O. Box 1450
Alexandria, VA 22313

Subject: ***Request for Comments on Intellectual Property Protection for Artificial Intelligence Innovation (Docket No. PTO-C-2019-0038)***

Dear Director Iancu:

The Association of American Publishers (AAP) appreciates this opportunity to provide its views in this request for comments on the impact of artificial intelligence ("AI") on intellectual property law and policy.

AAP is the national trade association of the U.S. book and journal publishing industry. AAP represents the leading book, journal, and education publishers in the United States on matters of law and policy, advocating for outcomes that incentivize the publication of creative expression, professional content, and learning solutions. The U.S. publishing industry supports an extensive network of American businesses and thousands of jobs. Publishers are an integral part of the U.S. copyright industries, which collectively added $1.3 trillion in annual value to the U.S. gross domestic product in 2017.[1] More so, publishing fuels American education and ingenuity, even more important in the 21st century than it was at the dawn of our democracy.

With the advent of smart devices, artificial intelligence has moved beyond a merely theoretical field to one with real world applications and marketplace implications. The availability of greater stores of digitized data and increased computational power are making the potential use of predictive AI in many areas such as transportation, health diagnoses, among others, a reality - advances that will have tremendous benefits for humankind.

Yet, what remains unclear is how policies will govern the development and application of AI. How will we appropriately recognize and compensate the owners of copyrighted works used to train AI systems? What ethical parameters will direct AI development? How will AI inform or impact private lives? There are numerous questions, and many potentially huge consequences. Such questions require equally serious government deliberations and robust public input.

Particularly in the area of intellectual property, and copyright especially, there is a danger that technology will far outpace or even ignore the legal framework. To date, policy discussions have focused on the question of ownership with respect to AI outputs (i.e., who should own what an AI program might

[1]Copyright Industries in the U.S. Economy: The 2018 Report, by Stephen Siwek, Economists Inc., https://iipa.org/files/uploads/2018/12/2018CpyrtRptFull.pdf

produce) but there has been little discussion of how the use of copyrighted works to "train AI" should be governed. This inquiry by USPTO, as it draws the focus on the impact of the development of AI technologies on intellectual property, particularly copyright, is timely and necessary.

Below AAP addresses the questions (most relevant to our members) posed in the request for comments, including whether an AI-created work can be accorded copyright protection and whether the current U.S. copyright law framework is sufficient to protect the rights of copyright owners whose works may be used to facilitate AI development.

At the outset, we note two things. First, while the query focuses on AI development, the request for comments does not provide a definition for "artificial intelligence," perhaps because there is no settled definition. Yet, while there is no agreed definition of AI, the varied definitions appear to come to rest around the notion of AI as data-driven *technologies* – from machine learning to natural language processing. The European Union in its 2018 Communication on Artificial Intelligence for Europe defined AI as referring to "*systems* that display intelligent behaviour, by analyzing their environment and taking actions – with some degree of autonomy – to achieve specific goals."[2] (Emphasis supplied.) As a "technology," AI cannot be severed from the human actors who design it or facilitate its ingestion of particular datasets necessary to instruct the AI program in addressing and finding a solution to whatever inquiry is presented. At its essence, AI is still *code* – written to accomplish the specific purpose or purposes intended by its human author or creator. While AI technology is becoming increasingly sophisticated as to itself be capable of writing code, it remains nonetheless a technology tool.

Second, it is axiomatic that the rights of creators and owners of copyrighted works should be respected and protected. While AI and its promise holds much allure, at the end of the day, AI is the sum of data-driven technologies designed and leveraged to accomplish specific tasks directed by humans. To train an AI algorithm or process, humans may feed it huge troves of data, which may in turn implicate valuable and proprietary copyrighted works. It is the position of AAP that the use of data that is expressed and embodied in copyrighted works should – in most cases - be permissioned and compensated, and where it is argued permission may not be necessary, the facts of use should be evaluated under copyright law.

Issues for Comment

1. **Should a work produced by an AI algorithm or process, without the involvement of a natural person contributing expressly to the resulting work, qualify as a work of authorship protectable under U.S. copyright law? Why or why not?**

AAP contends that a work created by an AI algorithm or process, entirely absent the involvement of a natural person in the creation of said work, does not qualify as a protectable work of authorship under U.S. copyright law.

Article I, Section 8, Clause 8 of the Constitution grants Congress the power "To promote the Progress of Science and useful Arts, by securing for limited Times to *Authors* … the exclusive Right to their respective Writings." U.S. law and the laws of other jurisdictions envision the "author" as a natural person,

[2] Artificial Intelligence for Europe, available at https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe

"who exercises subjective judgment in composing the work and who controls its execution.[3]" The output, therefore, of a truly independent *artificial* intelligence would not qualify for copyright protection under current law. Some have suggested that the work-for-hire doctrine serves as precedent for non-human authorship because it permits a corporation to be the author of certain works subject to certain conditions. The work for hire doctrine is, however, easily distinguished from AI output: while individuals may be pooling their collective accomplishments into one work-for-hire work, these individuals have still directed and executed all of the creativity so involved.

The output of an AI algorithm or process may, perhaps, qualify as "subject matter other than works," as so framed in other jurisdictions. Whether such a distinction is or will become necessary should be the subject of further consultations, following USPTO's assessment of the views received through this comment process.

2. **Assuming involvement by a natural person is or should be required, what kind of involvement would or should be sufficient so that the work qualifies for copyright protection? For example, should it be sufficient if a person (i) designed the AI algorithm or process that created the work; (ii) contributed to the design of the algorithm or process; (iii) choose data used by the algorithm for training or otherwise; (iv) caused the AI algorithm or process to be used to yield the work; or (v) engaged in some specific combination of the foregoing activities? Are there other contributions a person could make in a potentially copyrightable AI-generated work in order to be considered an "author?"**

Of the scenarios enumerated in the above question, AI-assisted creations in which a natural person either designed the AI algorithm or process or contributed to the design of the AI algorithm or process could conceivably qualify for copyright protection – assuming the human element brings the requisite modicum of creativity to be eligible for copyright protection. The designer/programmer will also have likely identified the data sets necessary to train the AI to accomplish the program's set tasks; in fact, it may be the case that the person choosing the data sets may only be following programing instructions or parameters as already defined by the designer/programmer, and not acting with any autonomy or injecting any degree of independent creativity into the process. Where an individual is only causing the AI algorithm or process to be used to produce the output for which the AI was programmed, that individual may merely be facilitating the execution of the program and not engaging in activity that contributes to the AI-assisted creation. These are all novel questions that require further consideration, the answers to which may evolve as AI systems evolve.

It is the human actor from whom originates the design and function of the AI program to whom authorship may be ascribed. Those that simply employ an AI program to facilitate accomplishment of the tasks for which the AI is designed are merely executing or carrying out the program instructions. (We do not speak here of ownership of the AI-generated work, which question may be addressed through or determined by the terms of the licensing or contractual agreement between the parties.)

3. **To the extent an AI algorithm or process learns its function(s) by ingesting large volumes of copyrighted material, does the existing statutory language (e.g., the fair use doctrine) and**

[3] Jane C. Ginsburg, "The Concept of Authorship in Comparative Copyright Law," 52 DePaul L. Rev. 1063 (2003), available at https://via.library.depaul.edu/law-review/vol52/iss4/3

related case law adequately address the legality of making such use? Should authors be recognized for this type of use of their works? If so, how?

Given the long success and adaptability of the U.S. copyright system to new technologies, we see no reason at this time as to why it cannot accommodate and inform the continued development of AI. This is the case not only because limitations such as the fair use doctrine may be relevant, as indicated in the above question. But more directly because the marketplace is sufficiently flexible and fast moving, easily able to propel and absorb new licensing schemes based on the exclusive rights afforded by copyright law. As AI technologies are continuing to advance, any changes to the copyright law framework – at this time – would be premature. In that vein, the industry strongly cautions against creating new exceptions, such as for text-and-data mining (TDM), for instance, to purportedly facilitate the use of data embodied in copyrighted works for purposes of AI training.

It is this industry's view that licensing solutions remain the best tool for facilitating AI development while also protecting the rights of authors, 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 – for instance, as regards the scope of use, authorship, and ownership of data sets and AI outputs.

Fair use may apply to certain uses of works to train AI, but that analysis will remain dependent on the facts attendant to each specific use of copyrighted works in developing or training an AI algorithm or process. Wholesale reliance on fair use would be troubling. Indeed, the uncertainty inherent in relying on only fair use is unlikely to be conducive to the development of a seamless system that protects the rights of authors and rights holders of the copyrighted materials used to "train" an AI algorithm or process while also facilitating its development.

Journal and educational publishers publish and curate large bodies of copyrighted works, which embody the datasets of greatest interest to entities wishing to engage in AI-facilitated research, to further machine learning and other AI enabled processes. Publishers are themselves also developers and users of AI technologies - to drive the analytics and educational services provided to their subscribers. Given the duality of publishers' engagement with AI technologies, it is important that clear policy guidelines be adopted to protect their investment in the publishing endeavor while also allowing publishers to continue to use AI technologies to enhance the services they provide to their customers.

4. **Are current laws for assigning liability for copyright infringement adequate to address a situation in which an AI process creates a work that infringes a copyrighted work?**

AI is currently viewed as a set of data-driven technologies, programmed by its human creator or designer to accomplish (a) specific task(s). If a work created by an AI process infringes a copyrighted work, the infringement can be ascribed to either the human programmer or the human operator of the AI, i.e., the human actor who ultimately either designed the AI algorithm or process or engaged the AI to produce the infringing output. Questions may arise as to whether the human actor is directly liable or merely contributed to the infringing conduct, in which case, the same analysis and assignment of liability under existing law is sufficient to address this matter. At its core, it is a question of which party has set the infringing conduct in motion and should, therefore, be held accountable.

5. Should an entity or entities other than a natural person, or company to which a natural person assigns a copyrighted work, be able to own the copyright on the AI work? For example: Should a company who trains the artificial intelligence process that creates the work be able to be an owner?

In the scenario outlined above, it is quite plausible for the entity using and training the AI process to own the output (or work), subject to third party rights. Questions regarding ownership of AI outputs can be more clearly defined via contractual or licensing arrangements between the parties. The scenario does not present a situation that is not or cannot be covered by existing law with respect to questions of ownership or assignment of rights. Perhaps analogous to the manner in which a company acquires copyright ownership over the works created by its employees, under a work for hire arrangement, an entity training and employing an AI algorithm or process to accomplish a task may similarly obtain ownership of the AI program's output if so agreed under a license or contract.

6. Are there other copyright issues that need to be addressed to promote the goals of copyright law in connection with the use of AI?

The publishing industry again wishes to underscore that the issues outlined by the USPTO require very careful and deliberate consideration. As developments in this area will continue to emerge rapidly, USPTO should continue to engage with rights holders and other agencies (both domestic and international) on the matter. The U.S. government should monitor efforts within international organizations to ensure that they not move faster than national governments in terms of policy setting in this area. In February 2019, the U.S. government released a statement[4] regarding America's leadership in artificial intelligence, yet nowhere in that piece is the importance of copyright stated. It is imperative that the U.S. government's policies continue to emphasize the importance of intellectual property rights, and its essential role in AI development.

9. How, if at all, does AI impact the need to protect databases or data sets? Are existing laws adequate to protect such data?

Although the United States considered legislating database protections at the time it implemented the WIPO Internet Treaties in the late nineties, such legislation did not move forward and we do not, therefore, have the kind of sui generis protection for databases or data sets available under the European framework[5]. U.S. law does, however, protect data collections or "compilations" where authorship "inheres in the way the compiler has selected or arranged the information."[6] As data, which may be embodied in copyrighted works, is critical to the training of AI, where such data exists in a compilation curated by a rightsholder, ingestion of the database to train the AI must be permissioned and/or compensated, or otherwise compliant with copyright law.

12. Are there any other AI-related issues pertinent to intellectual property rights (other than those related to patent rights) that the USPTO should examine?

---

[4] "Accelerating America's Leadership in Artificial Intelligence," February 11, 2019, available at https://www.whitehouse.gov/articles/accelerating-americas-leadership-in-artificial-intelligence/.
[5] EU Directive 96/9/EC on the legal protection of databases, available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31996L0009&from=EN.
[6] Id at 2.

As noted above, data is essential to AI training and development. The USPTO should examine how "data" is defined in various government instruments, including trade agreements, to ensure that they promote a consistent policy across U.S. government agencies. Care should be taken to ensure that "data" is not so broadly defined as to include copyrighted works, such as books, journal articles, and other creative works developed for and created by private sector rights holders.

Where "data" is defined in overbroad terms, it is particularly important to safeguard the intellectual property rights and other proprietary interests that may be swept up in a broad definition of "data." The Open Government Data Act (OGDA), which was enacted in 2019 as part of the Foundations for Evidence-Based Policymaking Act of 2018 (the Evidence Act),[7] is instructive in this regard. As explained in the accompanying report by the House Committee on Oversight and Government Reform, the OGDA makes clear that it would be inappropriate for the government to impose open license requirements on "data that the government uses and maintains but does not own." In addition to provisions that address instances involving contractual, privacy, security, business and other confidentiality considerations, the law contains provisions to ensure that agencies evaluate whether a data asset is "subject to intellectual property rights, including rights under titles 17 and 35, United States Code" to determine whether it can be made publicly available subject to an open license. The Evidence Act also contains a broad rule of construction to ensure that nothing in the law may be construed to override or limit intellectual property rights.

While data created by and for the government may properly be the subject of un-permissioned or uncompensated use for purposes of AI training, this does not extend to copyrighted works created and owned by private sector actors. Use of works protected by copyright law, unless covered by an exception to copyright law, should properly be the subject of licensing arrangements.

13. Are there any relevant policies or practices from intellectual property agencies or legal systems in other countries that may help inform USPTO's policies and practices regarding intellectual property rights (other than those related to patent rights)?

We would welcome USPTO using its expertise to monitor and share with U.S. stakeholders information pertaining to the policies and deliberations of other jurisdictions regarding IP and AI. Furthermore, we note that while AI systems may be poised to accomplish important advances, the U.S. and other governments should not sanction the ingestion of valuable intellectual property unconstrained.

The EU Directive on copyright in the Digital Single Market[8] (DSM) created a text-and-data mining (TDM) exception, in which it differentiated between TDM for commercial and for non-commercial purposes (specifically, for scientific research). Where the user (text-and-data miner) is a commercial or for-profit entity, the DSM requires the commercial user to obtain a license for "reproductions and extractions" of "works or other subject matter to which they have lawful access." Given the downstream commercial applications to which such a commercial data miner might apply the output of its mining of works, and the benefits that will likely accrue to the commercial actor, it is quite reasonable for the creator/owner of said works to be compensated for such use.

---

[7] Public Law No: 115-435 (01/14/2019)

[8] EU Directive on copyright in the Digital Single Market, available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L0790&from=EN.

In contrast, frameworks that provide too broad a carve out for un-permissioned and uncompensated uses, even where such use is of a commercial nature, may result in an eroding of rights accorded to rights holders that curate and own copyrighted works or compilations of copyrighted works. As noted above, any framework to promote AI development must not degrade or diminish intellectual property rights, including copyright.

Conclusion

We appreciate USPTO's careful entrée into the issues of AI and IP. While data is essential to the development of AI technologies, such data will in many instances be embodied in the copyright protected works of authors, publishers, and other copyright owners. The federally protected rights of these rights holders cannot be sacrificed to the altar of AI development. Government policies must be such that the pursuit of AI development does not unreasonably impinge on nor detract from the rights of creators and rights holders in whose works may be embodied data needed to train AI. The overall ecosystem must be balanced and rational.

While this early inquiry will be useful to informing the USPTO's initial consideration 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 the USPTO for the opportunity to respond to this request for comments and look forward to participating in further consultations on this important issue.

Sincerely,

M. Luisa Simpson
Senior Vice President, Global Policy