



605 5th Avenue S., Suite 400, Seattle, WA 98104 www.gettyimages.com

February 14, 2020

World Intellectual Property Organization

34, chemin des Colombettes

CH-1211 Geneva 20, Switzerland

Via email (ai2ip@wipo.int)

Re: Request for comments on DRAFT ISSUES PAPER ON INTELLECTUAL PROPERTY (IP) POLICY AND ARTIFICIAL INTELLIGENCE (AI)

Getty Images greatly appreciates the opportunity to submit the following in response to WIPO's request for comments on the draft prepared by the WIPO Secretariat identifying and discussing the issues arising for IP policy in relation to AI. We agree with the paper's assessment that AI is already having a significant impact on the creation and commercialization of cultural works and we support the development of policy that encourages innovation and creativity that is transparent, ethical and legal.

Getty Images is a leading source for visual content across the world, in that no other organization has the exact combination of creative imagery, vectors and video footage, combined with the comprehensive nature of our editorial imagery and video footage. We have a long history of managing high quality content, and our business model encourages the creation of artistic work by providing a system for lawful licensing and the monetization of content.

As technology such as AI and machine learning (ML) enable the visual arts to evolve, we are committed to protecting the intellectual property rights of our community of over 310,000 contributors as well as respecting the privacy and property rights of third parties. Although the potential applications of AI and ML are limitless and it is impossible to accurately predict what the future will hold, it is important to recognize that mature technologies available today, such as generative adversarial networks (GANs), require us to rethink the interaction between technology and the creative process. In the context of the visual arts, GANs have made possible AI tools that are capable of creating high-quality synthetic content for a low cost and at scale. As amazing as this is, such tools are not capable of true independent creativity. In order for AI tools to create new work, prior creative work must be used as training data. It is essential that any such work used as training data be tracked in an auditable record to ensure that it is handled in a lawful and respectful manner.

At Getty Images, we believe that now is the time to define a legal framework for AI and ML. The current lack of official IP guidance has already resulted in the creation of AI tools that have violated privacy and IP rights internationally. An effective legal framework can give all interested parties clarity and guidance creating commercial opportunities for human creators and enabling the responsible development of groundbreaking AI tools.

We are impressed by WIPO's commitment to engaging with these issues and are supportive of the proposed process for developing effective IP Policy in the context of AI. We hope that our answers help raise awareness of the issues that we believe are key and that our comments contribute to construction of the path forward. We feel strongly that solving the outstanding problems posed by these issues is key

to the future of the creative industries and the humans who fuel it. We would like to underline our support for urgent action as opposed to taking a “wait and see” approach. It is important that rules of the road are established as early as possible if the international creative industries and business at large are to have a sufficient level of legal certainty in order to, respectively, continue creating and investing in areas where AI has the potential to transform whole industries.

Comments on Issues

PATENTS

We have confined our comments on patents (Issues 1-5) to a single high-level issue that we believe is key to consider in order to effectively develop AI related policy across all types of intellectual property, that of the maintenance of auditable records of training data.

Issue: 4 Disclosure

10.(iv) How should data used to train an algorithm be treated for the purposes of disclosure? Should the data used to train an algorithm be disclosed or described in the patent application?

A legal requirement to track and disclose training data is critical to establishing a framework within which an assessment can be made whether AI tools are being developed and utilized in an ethical and legal manner. In addition to contributing to the goal of creating a public record of humanity’s technology, in the context of AI, tracking and disclosure obligations can help society identify when the creation or application of AI tools violates the rights of third parties. More specifically, policy that puts an obligation on the developers of AI tools and systems to keep auditable records of the training data that they use would enable trust and could promote the ethical and legal commercialization of the output of such systems. Therefore, we recommend that data used to train an algorithm should be required to be disclosed or described in a patent application. Furthermore, that auditable records should be required to be maintained for any use of the patented AI process.

Issue 5: General Policy Considerations for the Patent System

Getty Images does not think it is too early to consider these questions. While we certainly do not recommend the adoption of sweeping policy measures that could result in unforeseen negative impacts, we do believe that it is essential to develop and implement basic policy that is intended to protect third party rights. Without such protections, we will likely see continued disregard for IP, privacy and other rights, and third-party victims may be irrevocably harmed without access to recourse.

COPYRIGHT AND RELATED RIGHTS

Issue 6: Authorship and Ownership

The draft paper’s framing of the authorship and ownership issue correctly identifies that the questions posed go to the heart of the social purpose for which the copyright system exists. However, permitting copyright protection for AI-generated or “synthetic content” does not necessarily need to place equal value on human and machine creativity as the paper suggests. In many cases, the generation of synthetic content requires the input of prior creative work and the involvement of a human. Accordingly, the rights of the owners of such prior work as well as the level of human involvement need to be considered when developing policy positions. Getty Images believes that any such policy should

favor human creativity over machine creativity and appropriately recognize human contribution, including the contribution of protected human created IP used as underlying training data.

Specific responses to the questions posed under number 12:

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

If there is no human involvement in the creation of a synthetic work, the work should not be protectable. In the case where no human attribution is possible, permitting copyright protection would likely devalue creative human endeavor and have a negative impact on the significant contributions that creative industries make to global GDP. In the context of synthetically generated content created by GANs without human involvement, providing copyright protection could prompt irresponsible behavior that could displace the market for the work created by humans. To keep this from happening, we need to think of deploying novel techniques such as an obligation to track and license prior work used in data sets. This would promote and enable the generation of synthetic content in a manner that is respectful of intellectual property rights.

- (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 stated above, Getty Images believes that some level of human involvement/contribution be involved for an AI-generated work to be eligible for copyright. Although it may be possible for AI to produce creative work without the direct involvement of a human, it may be more helpful to view the technology as a tool that can help humans create new work. Viewed through this lens, it is prudent to rely on established IP law and policy to provide guidance on if the minimum level of creative mental input for copyrightability is met. If the output of AI meets the criteria for copyrightability and the contribution of a natural person is also sufficient, the copyright should vest in the natural person (or her employer) who made such contribution. We do not see a need for according a legal personality to an AI application itself.

In the context of the copyrightability of synthetically generated work created by GANs, one key issue that needs to be considered when determining who should have rights in any associated copyright is the role of underlying preexisting work that was used as training data. Policy makers must recognize that synthetically generated work is not possible without the use of preexisting work as training data. Accordingly, if the training data includes work protected by copyright, creative synthetic work generated by AI tools should be considered a “derivative work” of each of the pre-existing works that have been copied as part of the process and authorization to use such work should be required.

- (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?*

Yes, it is worth considering a sui generis system of protection for AI generated work. However, the benefits of creating such a system need to be weighed against the cost of confusion caused by adding such complications. Rather than treating such works differently

than human created works in the same class, it may be prudent to instead create unique standards that AI generated work needs to meet in order to qualify for protection. As discussed above, one such standard could be to condition protection on the tracking, licensing and disclosure of underlying work used as training data. Such measures can provide a safeguard against the misuse of such underlying work and enable the creation of mechanisms that can recognize and potentially compensate human creators for their contributions.

Issue 7: Infringement and Exceptions

Our vision at Getty Images is to have our images everywhere and that includes the use as ML training data, provided this does not prejudice the interests of the creators who we represent. We are supportive of AI innovation and have existing licensing in place that authorize the use of our data for such ML applications. However, it is key that the developers of AI tools who use protected data do so in a manner that is respectful of IP rights and the contributions of human creators. As emphasized in our initial statement above, the unauthorized use of creative work protected by copyright as training data is a primary concern for Getty Images and the human creators that we represent. Since AI, and in particular GANs, require significant data sets on which to train, often the best data sets are compilations of copyright protected works. While existing laws may in theory be adequate tools for assessing if a use is infringing, the threshold issue that needs to be quickly addressed is what obligations the developers of AI tools have to maintain auditable records. If such records are not maintained, it may be impossible to produce evidence that either supports a case for infringement or alternatively can be asserted to support a case for fair use or fair dealing exceptions.

Specific responses to the questions posed under number 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?*

It is the general position of Getty Images that the use of copyright protected work as training data requires preauthorization by the copyright owner. While we recognize that it is in the public's benefit to permit certain fair use / fair dealing exceptions of protected work (for example in some educational contexts or in a truly transformative manner), most current commercial AI use cases that rely on copyrighted work should require authorization. In the context of generative AI tools that are designed to create synthetic content, the importance of protecting the rights of underlying rights holders is especially high. The synthetic work derived from the copyrighted work used as training data not only depends on such work as the source of its "creativity," it has the potential to directly and unfairly compete with the market for that underlying work.

If we want to protect intellectual property rights, the law should require prior authorization in the context of using copyrighted content in a ML data set. Clearly, there are circumstances where authorization may not require a payment to the creator and the doctrines of fair use/ fair dealing provide workable guidance, e.g. in certain circumstances where the output of the data processing is pure scientific analysis and does not result in the creation of new content. However, regardless of the source of authorization, a heightened standard for copyright eligibility for AI generated synthetic content is warranted. Since it can otherwise be impossible to physically link synthetic content back to all of the underlying preexisting works, any natural person who wishes to receive and enforce copyright

protection for an eligible synthetic work should be obligated to track the body of preexisting works used for the training. The protection that this obligation would grant to copyright holders would justify any friction caused and may be handled by commercial contracts, possibly administered via collecting societies in respect of bodies, or parts of bodies, of preexisting works.

- (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?*

As stated above, while Getty Images strongly believes that authorization from the copyright owner is required prior to using copyrighted work, we also support and encourage innovation. Simply stated, we want such innovation to be conducted in an ethical manner that is respectful of IP and other rights. Ultimately, the goal of AI related policy should be to discourage fraud, misuse and other unfair practices. Requiring authorization for the use of copyright works in the commercial ML process promises to be an effective way to accomplish such goal.

We do not believe that a requirement to license data subsisting in copyright works for machine learning will hinder innovation in AI. Rather, it will provide a firm base upon which innovators may advance AI technologies with greater confidence. Accordingly, end users will be able to employ such technologies without the fear of unknown liability.¹

- (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 ?*

Yes, as further described our initial response to this issue above, we do support the concept of narrow exceptions for certain prescriptive non-commercial limited purposes. While there is certainly scope for more granular guidance as new technological processes and outputs emerge, international Copyright laws have viable and flexible mechanisms designed to provide guidance on what these exceptions should be. However, in the context of AI, in order to rely on such laws, the developers of AI tools must still be obligated to maintain auditable records that track what content was used as training data, even if it is clear that they can rely on an exception. Commonly, non-commercial “research” based projects that may qualify for an exception are converted or used for commercial purposes down the road. Without a requirement for record keeping in connection with excepted usages, such related commercial usages would never be able to be audited for potential infringement.

- (iv) *If the use of the data subsisting of 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?*

¹ For example, the explosive growth of the startup facial recognition company Clearview AI, has exposed end users to liability based on Clearview’s cavalier attitude toward using images as training data without obtaining authorization from rights holders. See <https://www.biometricupdate.com/202002/second-biometric-data-privacy-lawsuit-against-clearview-ai-also-includes-partner-supplying-police>.

Existing exception for text and data mining should be consulted when determining if the unauthorized use of copyrighted work should be considered non-infringing. Again, what is key in the context of AI is the obligation to preserve auditable records when protected data is used so that it is possible to adjudicate if an exception is warranted. If content is clearly of a type that is protected by copyright or an IP owner in good faith marks their content comprising the data as copyright protected, this should constitute sufficient notice to commercial data consumers that such preservation is necessary.

In connection with the two new exceptions for text and data mining included within the new EU Copyright Directive, we do feel that such exceptions can be improved and built upon, as they relate to AI and ML. For example, it should not be burdensome for a copyright holder to reserve their right in “an appropriate manner” if it does not wish to allow their copyright works to be used for text and data mining for AI and ML purposes under Article 4(3) of the Directive.

- (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?*

While well-crafted policy intervention may help facilitate the growth of the market for licensing in the context of AI, it is not “necessary.” As demonstrated by our successes entering commercial licensing deals with the developers of AI tools, it is clearly possible for commercial entities to enter into reasonable licensing agreements without policy intervention. At Getty Images, we want our content to be used everywhere and we believe that we can facilitate such use in a way that fairly compensates human creators.

- (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?*

As described above, developers of AI tools should be obligated to maintain auditable records that track what copyrighted work was used as training data. The disclosure of such records could be required by law in the context of an infringement dispute. For example, a defendant in an infringement suit could present the records to support an affirmative defense based on an exception. Without such an obligation, detection and enforcement could become an impossible task and rights holders could be irreparably harmed.

It is essential to consider policy solutions that encourage transparency in use. In the context of synthetically generated work created using GANs, regulators may want to consider limitations on the commercial exploitation of work that cannot be openly linked to an authorized data set. Such a framework would encourage a market for the legitimate sourcing of copyrighted training materials and, if crafted appropriately, can help ensure that the authors of underlying work be recognized and proportionately rewarded. Because this will likely entail micropayments being attributed to those individual copyright works used to train the AI algorithms, regulators may wish to also consider encouraging the establishing of collecting society schemes to plug any gaps unable to be filled by direct licensing.

Issue 8: Deep Fakes

In reaction to the considerable controversy surrounding deep fakes, it seems appropriate for policymakers to begin studying the harmful applications of synthetic creations. When doing so it is important to note that deep fakes pose the biggest threat in the “editorial” context. When synthetic content is used to intentionally spread disinformation, consumers of news can lose trust in the sources used for dissemination and editorial integrity is threatened. The result is a hybrid threat affecting human rights issues such as freedom of speech, defamation and slander, social cohesion, as well as national security.

However, the same underlying AI and ML technology—known as “generative adversarial networks” (GANs)—that manipulates visual content in order to make “editorial” deep fakes, is essentially the same as that used to generate, synthetic “creative” content that has potentially positive applications. The tools and apps to create bad deep fakes and good synthetic content alike are only increasing in number and ease of use, requiring common sense rules that can scale. Therefore, in developing rules around AI and ML and deep fakes, Getty Images recommends:

- Policy makers should stay neutral on and not ban underlying technology, while developing rules, regulations, or fines for unjust applications of AI and ML technologies.
- Assuming that our recommendations in relation to copyright protection above are followed, governments, early and publicly, should actively use copyright infringement in cooperation with copyright owners as justification in the take down of negative and harmful deep fakes. This will provide a playbook for action elsewhere and will direct synthetic content creators to obtain proper licenses for underlying content.
- Tools that can identify and verify “authentic” editorial content could be standardized and promoted. Metadata that can be used for authentication can also be used to track 3rd party rights and can be incorporated into systems designed to compensate such 3rd parties for the use of their work. For example, in the context of Art 17 of the EU Copyright Directive relating to user uploaded content, such metadata could be used to help social media platforms to meet their mandated obligations.

Issue 9: General Policy Issues

Privacy

Privacy is a central issue in policy discussion relating to technology, and the rise of AI and ML presents an entirely new set of challenges as well as other social and legal consequences. Although data that contains “private” information is not typically thought of as intellectual property, well designed IP policy should carefully consider the closely related privacy implications. New and emerging privacy regimes in Europe (GDPR), the United States (CCPA, BIPA), and elsewhere in the developing world have highlighted the importance of privacy to commercial actors, and regulators will want to determine what privacy rights are afforded to human subjects whose biometric and other private information is used to create AI tools and synthetic content.

Algorithms must use data to learn, correct, and become more effective. If that information is collected without proper privacy approvals, it is very difficult to have the algorithm “unlearn” it in order to satisfy a right to be forgotten or an underlying privacy violation. As AI learns, it accesses more data, makes more assumptions, and sends information more independently. The proactive nature of the technology makes data control more and more challenging.

Current privacy rules may be inadequate as policymakers are forced to rethink the right to broadcast or the right to resell biometric information of an individual or a crowd of people with or without permission. Content that is manipulated without the knowledge of an individual featured, flattering or unflattering, could raise questions of individual privacy, rights of publicity and/or defamation. It is essential that these issues be considered in the new age of synthetic content.

In developing recommendations on privacy for synthetic content, we recommend regulators articulate:

- The extent of privacy rights of human subjects included in both synthetic work and images contained in underlying data sets.
- That experimental application of emerging AI/ML tools for the creation of synthetic content are not an excuse for any privacy violations of subjects.
- That the burden to maintain the privacy rights of subjects incorporated into synthetic work rests on the creator of the synthetic work. Privacy needs to be addressed by such creators through disclosure and obtaining consent not through the unauthorized scraping of data or the reliance on data sets with known issues.²

Issue 10: Further Rights in Relation to Data

In response to the statement made in Paragraph 18, Getty Images strongly believes that data well suited for machine learning applications is economically valuable. In cases where such data is protected IP, that value needs to be recognized and rights holders need to be appropriately compensated for its use.

In response to the statement made in Paragraph 21, yes, the classical IP system does afford certain types of protection to data and policy makers should consider how to promote the values of this system in the context of AI and ML. Again, establishing the obligation to maintain auditable records will help policy makers apply established jurisprudence in these areas.

Issue 12: Capacity Building

Technologies such as AI and ML are often developed across jurisdictions and the market for AI generated synthetic content is global. It is therefore essential to take a global approach when considering policy especially if the goal is to reduce the technology gap in AI capacity. While it may be beneficial to promote capacity building in countries that currently lag behind through traditional economic incentives, it is important that general IP policy is consistent globally. Inconsistencies designed to promote capacity building (for example, relieving certain jurisdictions from an obligation to maintain auditable records) could have devastating effects on the global ecosystem for AI development and promote pockets of unscrupulous behavior. To counter this risk, we would recommend that certain minimum requirements be enshrined in new international copyright treaties, in particular requiring that developers and users of AI tools maintain auditable records of any training data used that is derived from copyright works.

² For example, the “MegaFace” dataset, which has been used extensively by AI developers, included millions of problematic images labelled as “creative commons” and has led to significant legal uncertainty. See <https://www.nytimes.com/interactive/2019/10/11/technology/flickr-facial-recognition.html>



605 5th Avenue S., Suite 400, Seattle, WA 98104 www.gettyimages.com

Issue 13: Accountability for Decisions in IP Administration

AI applications are well suited for tasks that require applying well defined rules and can be trained on abundant and diverse historical data. Accordingly, IP administration is clearly one such area where AI applications have the potential to excel. As with traditional, human assisted administration, mistakes will be made, and novel edge cases will always be presented. It is therefore important, to not fully automate the processes and include some form of human intervention. For example, the maintenance of a viable process to appeal decisions made solely by AI to a human arbitrator is key. To the extent that current appeal processes can be adopted to handle AI related appeals, legislative change seems unnecessary.

Respectfully submitted,

Paul Reinitz
Director, Advocacy and Legal Operations Counsel
Getty Images
605 5th Avenue S., Suite 400
Seattle, WA 98104
www.gettyimages.com