25 YEARS gettyimages | iStock

101 Bayham St, Camden Town, London NW1 0AG, United Kingdom

November 30, 2020

To: Intellectual Property Office

Via email: AIcallforviews@ipo.gov.uk

Getty Images response to IPO call for views: Artificial Intelligence and Intellectual Property – Copyright and related rights

Getty Images appreciates the opportunity to submit this response. At Getty Images we understand the potential value of the collection of technologies commonly referred to as "AI" and we embrace the change that advancement of such technologies will bring. While we fully support innovation in AI and Machine Learning (ML), it is clear that such innovation presents novel intellectual property considerations that require us to rethink the interaction between computers and the creative process. AI and ML can pose an existential threat to human creators and it is essential that innovators maintain high ethical standards when developing these new technologies. If such innovators can be encouraged to behave in a responsible, trustworthy and legal manner, there is real potential that the fruit of their labour will benefit humanity in many ways.

We have contributed to the responses to this call for views from the British Copyright Council and the national trade body for our sector, BAPLA. We are aligned with their submissions and this response concentrates on those areas where we feel we can add value from our own experiences and insights specific to our industry, the licensing of images and video clips.

Getty Images' Interest in AI

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.

Over the last few years, we have expanded our customer base to include companies in a wide variety of industries wishing to use our content in the development of AI tools and services.[1] Such customers know that they can trust the provenance of our content, leverage its unparalleled depth and breadth, and limit potential legal risk through licensing what to them is a critical input component. Our proven track record of licensing large volumes of visual content and associated metadata, delivered dynamically via our APIs, to customers throughout all the main continents, gives us a unique perspective on understanding how to effectively and ethically enable use of protected IP in the context of AI and ML.

[1] While we are unable to disclose the names of individual AI/ML license customers, the industries we work with are varied and represent a broad swath of AI/ML use cases that rely on protected IP. Non-exhaustive examples of industries within which these customers operate include: Technology (software and hardware), Agriculture, Automotive, Food Industry, Camera Manufacturing, Furniture Manufacturing, Beauty, Govt Institutions, Research and Non-Profit.

101 Bayham St, Camden Town, London NW1 0AG, United Kingdom

AI and the Creative Industries – Protection for Human Creators

As technology such as AI and ML enable the visual arts to evolve, we remain committed to protecting the intellectual property rights of our community of over 415,000 contributors as well as respecting the privacy and property rights of third party individuals. While the development of AI has many benefits, the risks that outputs such as artificially created synthetic content pose to third-party intellectual property created by humans are significant. We therefore commend the IPO for considering these issues together. Rapid developments in AI technologies and the demonstrated misuse of IP as a data source make it imperative to recognize IP rights as a subsection of any new AI legislation.

Synthetic Content and Generative Adversarial Networks

It is important to recognize that mature AI 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 new synthetic content, at scale and at negligible incremental cost. As amazing as this is, such tools are not capable of true independent creativity equivalent to human creative endeavours. In order for AI tools to create new visual works, prior creative work must be used as training data. It is therefore essential that any such work used as training data be tracked and an auditable record maintained, to ensure that it is handled in a lawful and respectful manner.

Time for Action is Now

At Getty Images, we believe that now is the time to define a legal framework for AI technologies. An effective legal framework for AI can give all interested parties clarity and guidance, creating commercial opportunities for human creators and enabling the responsible development of ground-breaking AI tools.

With Brexit, the UK has a unique opportunity to steal a march on other jurisdictions, taking note of the thinking within the EU institutions but being quicker to implement new laws and standards. We feel strongly that solving the outstanding problems posed by these issues is key to the future of the creative industries and to 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 are to have a sufficient level of legal certainty in order to continue creating and investing in areas where AI has the potential to transform whole industries.

RESPONSES TO QUESTIONS POSED

The use of copyright works and data by AI systems

1. Do you agree with the above description of how AI may use copyright works and databases, when infringement takes place and which exceptions apply? Are there other technical and legal aspects that need to be considered?

In general, yes, we agree with the IPO's description but, as the exceptions referenced do not mention machine learning or AI by name, it would be helpful to have further legislative clarity on the narrow circumstances in which they may apply in this specific context, including clarification on when they can never apply, e.g. we would argue they should never be able to apply where a potential output of the ML or AI process could be a competitive substitute for the input data (such as where images are used as data in a GANS system to create derivative synthetic content).

Other legal aspects to be considered are: (i) the Séjourné report on "Intellectual property rights for the development of artificial intelligence technologies" (2020/2015 (INI)), which was adopted by the European Parliament on 20th October 2020[2]; and (ii) to the extent not covered by other parts of the IPO consultation, the privacy implications of biometric data of individuals depicted in photos and other copyright works being used by AI systems such as facial recognition applications.

Pertinent points recommended by the Séjourné report include:

"...Stresses the key importance of balanced IPR protection in relation to AI technologies, and of the multidimensional nature of such protection, and, at the same time, stresses the importance of ensuring a high level of protection of IPRs, of creating legal certainty and of building the trust needed to encourage investment in these technologies and ensure their long-term viability and use by consumers;...(article 6)"

"...Recommends that priority be given to assessment by sector and type of IPR implications of AI technologies; considers that such an approach should take into account, for example, the degree of human intervention, the autonomy of AI, the importance of the role and the origin of the data and copyright-protected material used and the possible involvement of other relevant factors; (article 9)"

"...considers that non-personal auditable records of data used throughout the life cycles of AI-enabled technologies in compliance with data protection rules could facilitate the tracing of the use of copyright-protected works and thereby better protect right-holders and contribute to the protection of privacy (article 17); ..."

2. Is there a need for greater clarity about who is liable when an AI infringes copyright?

We recognise that due to the "black box" nature of AI systems, determining infringement of a specific work after the event can be a challenge. We believe strongly that licensing solutions are an ethical imperative and the main way of avoiding risk of infringement when protected IP is used in connection with AI and ML. In this regard, we would argue that there should be a statutory requirement for developers of AI across all sectors to retain auditable records that detail what training, testing and validation data was used in AI development and the ML process, especially if protected IP is involved. Then, where input data contains copyright works, questions about whether an exception can be relied upon could be determined subsequently. A legal requirement to maintain auditable records will also help instil trust in AI systems, by enabling developers and operators of those systems to demonstrate that they have used "good data" that is less likely to lead to discriminatory or biased outcomes.

3. Is there a need to clarify existing exceptions, to create new ones, or to promote licensing, in order to support the use of copyright works by AI systems? Please provide any evidence to justify this.

Licensing: Promotion of the licensing of copyright works used in the development and training of AI systems and ML is of paramount importance. Images together with associated metadata are incredibly rich sources of training data and if the human creators of those images are to share in the value generated by this new technology, it is critical they be licensed at the outset.

---

[2] Announcement on the European Parliament's website: https://www.europarl.europa.eu/news/en/press-room/20201016IPR89544/parliament-leads-the-way-on-first-set-of-eu-rules-for-artificial-intelligence and text of adopted report: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0277_EN.html

As noted in our introductory comments, Getty Images already licenses high-volumes of images and video clips, together with their associated metadata, for use in connection with AI technologies to scores of customers across various industries globally. These customers recognise the importance of building their AI and ML technologies on a solid foundation by ensuring that rights to their data inputs are legally acquired, in the same way that other raw materials are acquired. Yes, this has a cost but it is not an impediment. The experience of Getty Images demonstrates this. The way that Getty Images approaches this space is to include an explicit restriction in its standard form license agreement prohibiting any use of content and metadata for any ML or AI purpose, then implements bespoke license agreements with customers for carefully defined AI or ML related purposes with appropriate additional safeguards included and ensuring that all licensed content is allocated an equitable share of licensing fees.

When content is licensed by Getty Images for these purposes, micro-royalties are payable to its contributors and content partners. This is an easily scalable model for all types of copyright content. APIs can be used to deliver content and report usage, with this information being seamlessly linked to royalty systems so that royalty payments automatically flow back to copyright holders. Whilst the value of individual royalties on single items of content under each licensing deal might be exceptionally low, the aggregate value over time can be meaningful.

Whilst we don't advocate major changes to the copyright framework, two ways that the UK government could promote licensing of copyright works in the context of AI and ML as part of any new AI legislation are:

(1) to require all developers of AI technology to maintain auditable records of all data used to develop and train their AI tools, including identifying any data that contains or is derived from copyright works and the source of such copyright works; and

(2) develop and actively promote record keeping standards that may be used in fulfilment of this requirement to maintain auditable records. For example, standards for record keeping associated with financial accounting have been in place since ancient times. If data used in connection with AI and ML is viewed as a type of currency, independent records that track its flow can be an essential tool as the AI ecosystem continues to evolve.

Exceptions: We certainly do not see a need to create any new exceptions. However, as noted in our response to Q.1 above, a narrowing and clarification of the parameters of the existing TDM and temporary copying exceptions in the specific context of AI and ML use is desirable.

At the very least, in addition to the existing caveats contained within these exceptions, it would be helpful if a fair dealing test were to apply to such exceptions.

Another helpful clarification would be for it to be explicitly acknowledged that data obtained from crawling websites without explicit authorisation does not constitute "lawful access" under Section 29A(1) of the TDM exception, especially if the website includes user T&C that explicitly prohibits such data mining or extraction.

4. Is there a need to provide additional protection for copyright or database owners whose works are used by AI systems? Please provide any evidence to justify this.

As described in our responses to questions 2 and 3 above, a new mandatory obligation to maintain auditable records that detail what and how work protected by copyright is used in the context of AI and

ML is essential. Such an obligation would help to solve the "black box" problem and would promote ethical and responsible development by removing the temptation to obscure how protected IP is used.

Protecting works generated by AI

5. Should content generated by AI be eligible for protection by copyright or related rights?

In order to answer this question, it needs to be clarified what is mean by "content generated by AI". For content generated exclusively by AI without any human creative intervention, we think the answer should be "no, such works should not be eligible for protection by copyright or related rights". This is because no economic incentive is required in these circumstances and, for the reasons given by the BCC in its submission, the implications for human creative endeavour could be devastating (i.e. human creators unable to compete with AI content created at scale and at negligible cost, plus also being deterred from creating new content for fear of infringing rights in AI generated content).

However, where AI is used as a tool to aid human creative endeavour resulting in original works according to established copyright principles, i.e. author's own intellectual creation reflecting their personality, per EU standards, then we would argue that such works could, in very limited circumstances, be afforded protection by copyright, with the threshold for originality and human creative input in this context being set at an exceptionally high level compared to other methods of creation. New legislation could perhaps provide guidance on how such tests should be interpreted where AI is used as a tool.

Also, a pre-condition to any new, original "AI assisted" works being afforded copyright protection should be that if any copyright works have been used as input data at any stage in the development or deployment of the AI system, then copyright in the output should not arise unless such underlying works have been licensed in advance, irrespective of whether an exception might apply. Moreover, the use of underlying works could give rise to questions around the creation of derivative works and/or co-ownership of the resulting works.

This all leaves the question of what to do about section 9(3) of the CDPA regarding computer generated works? Considering that this is an outlier globally, that there has been only very limited reliance on this provision in the English courts and that it has the potential to confuse the position with regard to works created by or with the assistance of AI technology, we query whether it should be repealed in favour of the above suggested guidance regarding new "AI assisted" works?

6. If so, what form should this protection take, who should benefit from it, and how long should it last?

In the event that AI is really used as a tool that enables a human creator to create a protectable work and the pre-conditions describe above are met, the form of protection does not necessarily need to differ from similar protectable work created using traditional tools and techniques. Creating a new category of AI created works would most likely lead to unnecessary confusion in the market.

7. Do other issues need to be considered in relation to content produced by AI systems?

With regard to the question of enforceability of copyright, current laws for assigning liability for copyright infringement are generally adequate to address a situation in which an AI process creates a work that infringes a copyrighted work. Liability should generally be assigned to a natural person who

either has control of the AI or who benefits the infringing work. While in theory such assignment can be consistent with current laws, in practice, this issue again highlights the need for an obligation to create and preserve auditable records. For example, if a developer uses copyrighted content as training data in a GAN without authorization and markets a resulting tool that enables the creation of synthetic content, liability should be assigned to him or her. In addition, if a user of such an unauthorized tool then creates synthetic content with it, such content would also be unauthorized and there should be liability assigned to that user as well.

Copyright protection for AI software

8. Does copyright provide adequate protection for software which implements AI?

In general, software which implements AI should be able to benefit from the same level of protection that other software is entitled to. The question becomes more complicated when AI is effectively creating the software and the level of human involvement is minimal. In such a circumstance, the principles outlined in our answer to question 5 above re content generated by AI should be applied.

9. Does copyright or copyright licensing create any unreasonable obstacles to the use of AI software?

On the contrary, from a policy perspective we feel that there should in fact be obstacles that keep copyrighted content from being used without authorization. Any such obstacles are not unreasonable and are in fact necessary to protect human creators from suffering irreparable damage.

To recap, it is the general position of Getty Images that the use of copyright protected work as training data requires preauthorization by the copyright owner, except only where an AI developer can demonstrate that its use is covered by an exception to copyright infringement. While we recognize that it is in the public's benefit to permit certain excepted uses of protected work (for example for pure scientific research purposes), 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. In order to protect intellectual property rights, the law should clearly require prior authorization in the context of using copyrighted content in a ML data set.