gettyimages®

January 10, 2020

Director of the U.S. Patent and Trademark Office

P.O. Box 1450

Alexandria, VA 22313-1450

Via email (AIPartnership@uspto.gov)

Re: Request for Comments on Intellectual Property Protection for Artificial Intelligence Innovation

Getty Images appreciates the opportunity to submit the following in response to the U.S. Patent and Trademark Office's (USPTO) Request for Comments on Intellectual Property Protection for Artificial Intelligence Innovation published in the Federal Register on October 30, 2019.

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 artificial intelligence (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 250,000 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 independent creativity and prior creative work must be used as training data. Accordingly, it is essential that any protected IP used as training data be handled in a lawful and respectful manner.

At Getty Images, we believe that <u>now</u> is the time to define a legal framework for AI and ML. Such a framework can give all interested parties clarity and guidance on how prior works be tracked and IP respected in the context of AI. If the copyright community can accomplish this goal, we can create commercial opportunities for human creators and enable the responsible development of groundbreaking AI tools. We are encouraged that the USPTO is interested in this area and feel that the questions posed are an excellent way to progress the public dialog. We hope that our answers help raise awareness of these issues and help find a 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's important that rules of the road are established as early as possible if the creative industries and US 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.

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

Although work produced by an AI algorithm or process often requires the input of prior creative work and the involvement of a human, if there is no human involvement in the creation of the 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 US GDP.[1] 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 prior work used in data sets enabling AI synthetic content to be generated in a manner that is respectful of intellectual property rights.

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) chose 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"?

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 title 17 to provide guidance on if the minimum level of creative mental input for copyrightability is met. Accordingly, the factors outlined in the question posed should all be considered on a case by case basis per established law. It's conceivable that each of the referenced types of human involvement could, in certain limited circumstances, be sufficient to qualify for copyright protection and published guidance from the USPTO and Copyright Office would be beneficial. However, it is key that any such guidance be adaptable to technology as it continues to advance and state clearly that some level of original and creative human input is required.

In the context of the copyrightability of synthetically generated work created by GANs, it is also important to consider the underlying preexisting work that was used as training data. 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" under title 17 of each of the pre-existing works that have been copied as part of the process.

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

---

[1]Per the US Department of Commerce, the US arts and cultural sector contributed over $800 billion to US GDP in 2016 (https://www.arts.gov/news/2019/latest-data-shows-increase-us-economy-arts-and-cultural-sector)

authorization may not require a payment to the creator and the fair use doctrine provides 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 and disclose 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.

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 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?

The short answers to the first two of these questions is: "no", existing statue and case law does not entirely address this situation adequately; and "yes" authors should most definitely be recognized (and rewarded) for such use of their works.

Whilst, in theory, existing statutory language and related case law could be effectively applied to address the legality of ingesting large volumes of copyrighted material in the context of AI, new legislation specifically aimed at addressing this new and evolving use of copyright material would be incredibly helpful and reduce the risk of future administrations needing to unpick caselaw developed by judiciary not always expert in these new technologies and being forced to deal with a disproportionate weight of argument put forward by deep-pocketed litigants who might have no interest in maintaining strong protection for intellectual property. The existing law does not need dismantling, it does provide a workable framework for protecting the interests of copyright owners while permitting certain "fair use", but new statutory guardrails are needed to ensure existing law is sensibly and fairly adapted to consider unique attributes associated with the use of large volumes of copyright work. In this regard, it is essential that the ingestion of any volume of copyrighted material in connection with AI learning is not considered a "transformative" fair use by default. Moreover, it is important that certain commercial use of the output of AI should never be allowed to be created reliant on fair use. One such output is synthetic content as this can be capable of acting as a substitute for copyright works ingested as part of the training data.

The reality of how copyrighted material is used in the context of AI can make it practically very difficult to determine, after the fact, what material was used. Like the tracking obligation proposed in the discussion of copyrightability in Question 2 above, it is essential to consider solutions that encourage transparency in use. Without such transparency, determining liability for the unauthorized use of copyrighted content will be incredibly difficult. 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.

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?

In the circumstance outlined in the question, 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 a tracking obligation. 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 that 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.

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?

If the output of AI meets the criteria for copyrightability (as discussed in our answer to question 2 above) and the contribution of a natural person employed by an entity is also sufficient, first ownership by such entity may be appropriate. However, as we comment in response to Question 1, there does need to be a requisite degree of human involvement.

With the help of guidance by regulators, a robust market for authorized use of copyrighted content in connection with AI can develop and the questions of ownership of output will most likely be determined by direct licenses and other commercial agreements.

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 promotion of the goals of copyright has always included a balance of interests and advances in technology have consistently made it difficult to obtain that balance. AI and ML are technologies that can be used to generate the type of creative content historically produced by human creators and, as a result, they are technologies that can completely change how creative content is produced. To protect the rights of copyright owners and make sure that their property is respected, we need to proactively define how AI effects the legal framework. While Getty Images believes that issues addressed in Questions 1-5 are of primary importance to address now, we acknowledge that they are only the tip of the iceberg when it comes to IP and AI. We welcome the opportunity to participate in any further dialog.

Questions 7-8

While Getty Images respects and understands the value of trademark, our comments here are focused on AI and Copyright.

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

As described above, 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 be adequate in theory, they may not be practically suited to the task at hand if there is no obligation to keep records of what data was used. Our vision at Getty Images is to have our images everywhere and that includes the

use in training data. However, such inclusion must be respectful of IP rights and the contributions of human creators.

Questions 9-12

Getty Images declines the opportunity to respond to these questions which our outside the area of our immediate interests.

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)?

Since 1988 when the US recognized the need for global IP cooperation with the Berne Convention Implementation Act, the internet has made the importance of consistency in global intellectual property laws even greater. Technologies such as AI and ML are often developed across jurisdictions and the market for AI generated synthetic content is global. We applaud the USPTO for taking a global approach and specifically urge the collaboration with the European Union as it has demonstrated solid leadership in reaction to technological change with its recent Copyright Directive. That said, the two new exceptions for text and data mining included within the new EU Copyright Directive can be improved and built upon, as they relate to such uses for AI purposes, e.g. by adopting our above various suggestions.
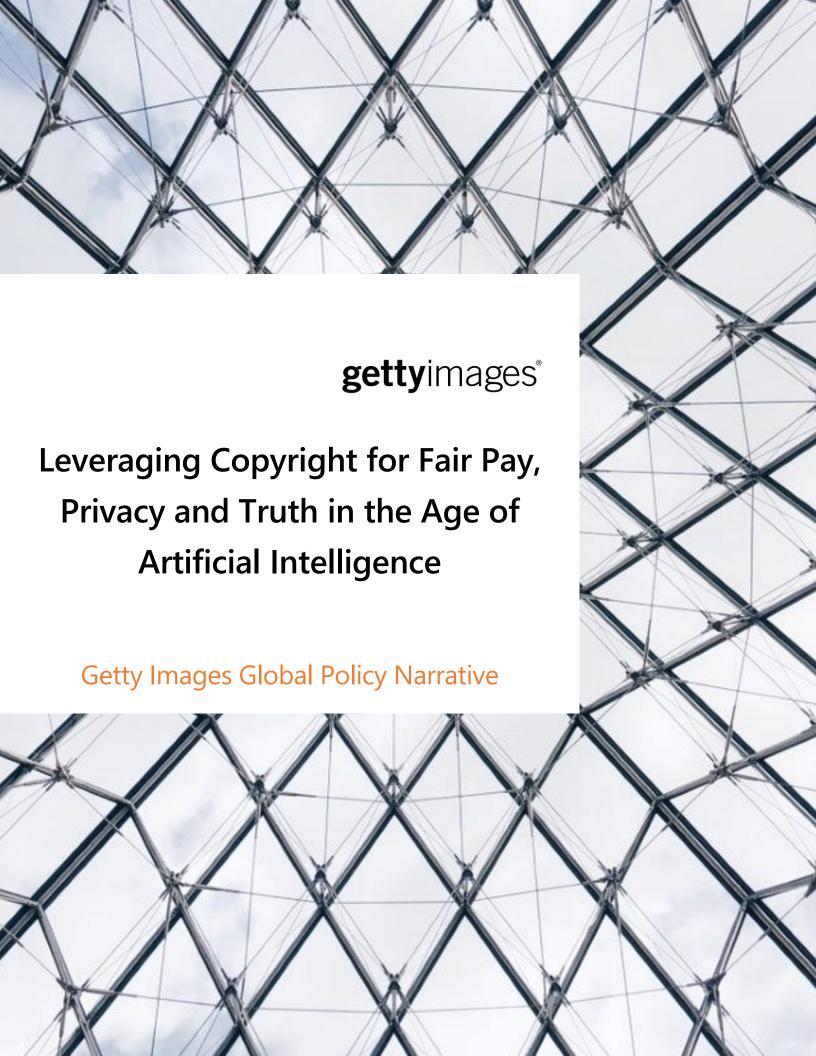

Respectfully submitted,

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

Attachment:

Getty Images Synthetic Content Policy Narrative

In addition to our responses to the questions posed by the USPTO, Getty Images would like to share our attached synthetic content policy narrative. The narrative goes into further depth on our thoughts related to copyright and AI as well as the issues of privacy and fraudulent misuse that we feel are intertwined and particularly pertinent to addressing concerns surrounding "deepfakes" and the like.

gettyimages®

Leveraging Copyright for Fair Pay, Privacy and Truth in the Age of Artificial Intelligence

Getty Images Global Policy Narrative

Contents

Introduction

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. Our vast and unmatched database includes two main categories of images, "creative" and "editorial" content. The "creative" category includes legally cleared content that can be used for a wide variety of commercial uses. Customers who license "creative" content can rely on us to manage the complicated assortment of rights associated with what is depicted in an image such as the rights of individuals, properties and brands and can be assured that the interests of rightsholders are respected. In addition, our "editorial" category includes "news/editorial" coverage that meets high standards of editorial integrity. Our "editorial" content is internationally trusted as an authentic representation of newsworthy events and is licensed primarily for editorial purposes with additional clearance typically required for commercial usage such as promotion of publicity. As technology such as artificial intelligence (AI) and machine learning (ML) enable both the "creative" and "editorial" disciplines to evolve, we are committed to protecting the intellectual property rights of our community of over 250,000 contributors as well as respecting the privacy and property rights of third parties. We believe that establishing a legal framework for AI and ML that is respectful of IP and privacy issues will create commercial opportunities for human artists and enable synthetic content to be used commercially in a legal and responsible manner.

Policy Narrative

Getty Images is a leader in facilitating the fair and efficient commercial exploitation of copyrighted content. We monetize content utilizing a platform that respects the rights of creators and gives end users comfort that they are meeting their legal obligations. As AI/ML create new commercial possibilities for synthetic content, it is essential that the ideas of lawful access, creators' rights, privacy/property rights and fair renumeration are considered by industry participants and policy makers. The ideas of lawful access, licensed content and fair remuneration that Getty Images has pioneered in the digital space should be expanded to the business of synthetic imagery created with (AI/ML) tools.

As Getty Images further develops its business model to align with changing technology, our current system of licensing copyrighted content is the basis upon which we will build our synthetic image business. However, we need support from policymakers to ensure the Getty Images business model can continue to provide lawful and licensed content to the market. Accordingly, the law must evolve with technology. We believe that clearly established rules of the road will give synthetic image creators the regulatory certainty to both experiment and scale globally.

It is essential to recognize that AI tools capable of creating synthetic content must be trained on existing data/imagery and the artists who own the intellectual property rights in such underlying imagery deserve to be compensated if their work is used for such training. In addition, the privacy and property rights of individuals and brands depicted in the underlying imagery need to be carefully considered. Licensing the original content used as training data that underlays a synthetic image provides fair pay for rightsholders, respects privacy and

property rights, and enables a legal audit trail for future legal issues concerning privacy, rights of publicity or defamation.

A Positive Vision for Synthetic Content

Getty Images' vision for synthetic content is to enable the rise of an entirely new artistic genre and the exponential creation of new content, all underpinned by lawful, licensed prior work. Getty Images is excited about the multitude of potential uses of responsibly created synthetic content. In the context of commercial "creative" content, such imagery can have powerful effects across society, including social justice. For example, Getty Images is currently leading conversations around re-picturing gender, disability, sexual orientation, mental health, an aging population and diversity. One example is our *Project: #ShowUs* initiative. Our image partners who have joined us on our mission to "Move the World with Images" include leading global commercial and not-for-profit brands. Synthetic images derived from unbiased training sets can help us illustrate not only what is, but what "can be" in a powerful way.

While we are excited by the prospects of synthetic imagery in the Getty Images family, we recognize that as with all technology, synthetic imagery presents opportunities for both good and bad. We believe the Getty Images business model presents an opportunity to ensure the best from synthetic images. However, we must identify and prepare meaningful policy not only for the positive actors, but also in anticipation of harmful ones.

> As a purveyor of both editorial and creative content, Getty Images is uniquely placed to balance the assets that new technologies provide with the need to ensure that photographers can continue to practice their craft and receive fair compensation. Getty Images' licensing system already protects those rightsholders and remunerates them.

Challenges & Opportunities for Policy-maker Consideration

Copyright

While we see an increasing level of global policy-maker attention given to the implications of AI/ML and the ethical use of AI/ML, one critical area seems consistently absent from discussion — namely, copyright protection and the rights of human artists. Given that copyright industries are struggling with illegal piracy in the age of digital distribution and given the high profile that world leaders have placed on fair remuneration to rightsholders whose content is used by online platforms, it is surprising to see this topic not fully considered in the AI/ML ethics debate, especially with regard to synthetic content.

While there has been some discussion of the applicability of copyright to synthetic images, there has been little thought expressed in connection with underlying content, protected by copyright that is used as training data and how the associated rights should be tracked. It is essential that these key aspects of copyright protection be considered by policymakers.

Underlying content

Copyright
Protection
& Tracking

Synthetic image creator

1. Underlying Copyrighted Images

The first consideration for policymakers is with the underlying copyrighted images. A key element of artificial intelligence, particularly in the development of synthetic images, is legal access to and use of high-quality underlying content for training purposes.

Not all creators of synthetic images, however, are mindful of this obligation or relevant copyright laws. As the technology evolves, it will be important to clarify these rights. Policymakers need to clarify that intellectual property rights must be respected in the context of AI/ML and provide clear guidance on the circumstances in which rightsholders must receive remuneration if their work is used as training data and that permission must always be obtained from such rightsholders if the output is synthetic content.

In addition, policymakers need to consider the necessity of imposing tracking requirements that will enable fair renumeration to the owners of intellectual property that is used for training purposes. Currently, there are no established standards or obligations that require the developers of synthetic image tools to track or identify what copyrighted content is included in the training sets. While this practice may be acceptable in the context of non-commercial research, as the market for synthetic images matures and commercial usage becomes common, such records will be key to establish legal marketability and ensure the appropriate respect for underlying rights.

2. Synthetic Image Owner

The second element for copyright consideration is the synthetic image itself. As with so many other creative endeavors, there will be many layers of creativity flowing into a synthetic image, and the creative artists involved in all layers of the activity should have the ability to copyright their work.

Copyright rules must be present to support that vision. This raises a wealth of new regulatory and policy challenges in jurisdictions globally, including who owns any rights to synthetically created images that are deemed to arise. Getty Images believes it is important to begin a dialogue on synthetic image ownership and underlying rights, including what elements are copyrightable.

In developing national or international licensing rules, Getty Images would like to put forward the following recommendations:

- Explicitly recognize images as a category in all new regulation on synthetic content;

- Require transparent record keeping that clearly tracks what copyrighted content is used as training data.

- Where necessary, modify existing exceptions to copyright to clearly require prior authorization from a rights owner in order to use their copyright protected work as training data in connection with the development of synthetic content and synthetic content creation tools, as well as in connection with other commercial uses. More specifically, recognize, in the context of a "fair use" or "fair dealing" analysis, that the output of generative models/algorithms which can produce synthetic content at scale at almost no cost, may unfairly and directly compete in the market with underlying training data and prejudice human creators.

- Where applicable and when a threshold level of human input is involved, introduce copyright protection for synthetic content.

- Apply the same principles on licensing contained in "traditional" copyright law when regulating synthetically produced content.

> Because Getty Images maintains a database of licensed content, unlike other actors, it has clear knowledge and control of the data it accesses. Getty Images' search engine tools enables researchers to find and confirm whether an image in question resides in our collection. This is not the case for other actors.

Privacy

Privacy is a central issue in technology discussions, and the rise of AI/ML presents an entirely new set of challenges as well as other social and legal consequences. Considering new and emerging privacy regimes in Europe (GDPR), the United States (CCPA, BIPA), and the developing world, regulators will want to determine what privacy rights are afforded to human subjects whose biometric and other private information is used in the process of creating synthetic imagery.

Algorithms must gather 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.

As an increasing number of images are posted online, it is increasingly challenging to seek permission from individuals who may be included somewhere in that image, or indeed may not even be a real person. Privacy rules may be inadequate as policymakers are forced to rethink the right to broadcast or right to resell the image of an individual or a crowd of people with or without permission. Images that are 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 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 reliance on data sets with known issues.[1]

Harmful Content, Editorial Deep Fakes, and Law Enforcement Concerns

Policymakers are beginning to study the harmful applications of synthetic creations, commonly known as "deep fakes." Deep fakes are a form of harmful "editorial" synthetic content whereby the individual subject or event

---

[1] 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

someone is watching may not exist, never happened, was never said, contradicts empirical facts, shames and disparages, or intentionally spreads disinformation. They are most often in video or image format and are a threat to editorial integrity if held out to be authentic. 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/ML technology—known as "generative adversarial networks" (GAN)—that manipulates visual content in order to make "editorial" deep fakes, is essentially the same as that used to generate legitimate, synthetic "creative" content. 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/ML and deep fakes, Getty Images recommends:

- Governments should stay neutral on and not ban underlying technology, while developing rules, regulations, or fines for unjust applications of AI/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 should be standardized and promoted.

As a leading visual communications expert and content producer, Getty Images upholds the responsibility to ensure accurate and authentic visual representation. Getty Images' award-winning editorial content includes trusted, watermarked and time-stamped first-hand photo and video footage of historical events and can provide the tools to authenticate and fight online disinformation. We take our responsibility to deliver authentic content seriously and our content license agreement restricts any alteration of editorial content that would compromise the editorial integrity of the image. Additionally, copyright law enables content creators to assert rights to the image or content being manipulated and thus may offer a legal recourse for law enforcement and victims against propogandists and purveyors of deep fakes and disinformation.

Conclusion

Synthetic content developed from licensed underlying work and transformed by AI/ML promises a variety of new benefits to society. It is also disruptive. Like any technology, synthetic content can be used by human actors for either positive or negative outcomes. Informed by lessons learned in two major markets on either side of the Atlantic, Getty Images is in a unique position at the crossroads of the arts and technology to offer new tools and perspectives to policymakers grappling for answers that are both

innovative and thoughtful. We are ready to work with decision-makers to create solutions that provide fair pay and renumeration, privacy and dignity, justice and truth.

> Getty Images is a positive actor who seeks to create a lawful, fairly remunerated market for synthetic content. Our licensing model also enables legal and regulatory auditability to help define and track the inputs of the aggregated final synthetic product.