



October 30, 2023

Suzanne V. Wilson,
*General Counsel and
Associate Register of Copyrights.*

Maria Strong,
Associate Register of Copyrights and Director of Policy and International Affairs.

U.S. Copyright Office
Library of Congress

Re: [NOI Federal Register Version, 88 Fed. Reg. 59,942 \(Aug. 30, 2023\)](#)

COMMENTS OF DIGITAL MEDIA LICENSING ASSOCIATION

Dear Ms. Wilson and Ms. Strong:

On behalf of the Digital Media Licensing Association (“DMLA”), we are submitting the Notice of Inquiry on the subject of Artificial Intelligence and Copyright initially published on August 30, 2023 (“NOI”).

DMLA is the leading trade association representing the interests of entities who primarily license still and motion images to publishers, the media, advertisers, designers and others. Its members represent the interests of thousands of photographers/videographers and the copyrights in millions of still and motion images by aggregating images and making them readily searchable and available for licensing online or via apps. DMLA supports a robust copyright regime that supports a thriving licensing industry, whether it's the licensing of media content, the metadata that is associated with digital content, or the licensing of a dataset of curated content that is available to entities for use as training models. DMLA supports the potential and opportunity that generative AI models offer but wants to insure that the models respect copyright, that training content is lawfully obtained, that there is transparency in the content that is used to train the models, including that records are maintained and accessible in order to identify content that is used in training, and that the creative community continues to receive benefit if their works are used in the creation of generative.

Cowan, DeBaets, Abrahams & Sheppard LLP

41 Madison Avenue
38th Floor
New York, NY 10010

9454 Wilshire Blvd
Suite 901
Beverly Hills, CA 90212

cdas.com

DMLA is a member of the Copyright Alliance and supports in general its submission to this NOI except for any differences described in this submission. DMLA's comments primarily address issues that are relevant to the visual content licensing industry in general, and not all questions are responded to. Individual members of DMLA may be submitting reports on their own behalf and the comments of DMLA do not necessarily represent the views of any individual member.

General Questions:

1. *As described above, generative AI systems have the ability to produce material that would be copyrightable if it were created by a human author. What are your views on generative AI systems potential benefits and risks of this technology? How is the use of this technology currently affecting or likely to affect creators, copyright owners, technology developers, researchers, and the public?*

DMLA and its members are in a unique position to view the potential as well as the risks of text to image generative AI systems to creators, copyright owners, technology developers and the public. DMLA members have been aggregating large libraries of digital visual and audiovisual content for decades on behalf of creators and have developed sophisticated platforms to make their content readily licensable to the media and others with a shared revenue model. In doing so DMLA members are involved with creators, technology companies as well as the public.

The current proliferation of text-to-image generative AI models is a significant development as many models use members' content for training models that can be used to generate visual content that can compete with these same visual libraries. Scraped training data from the internet instead of licensing content for data sets contravenes copyright law and allows generative AI companies to use this visual and audiovisual content without proper compensation. The practice of treating the Internet as a sea of free resources to take without adhering to longstanding copyright regimes throughout the world threatens the viability and future of the media licensing industries.

There is also risk to the public if generative AI is built on unlicensed content without any controls on the output. The quality of AI generated content is improving at a fast pace and the ability to create synthetic or “fake” visual and audio-visual works can erode the trust in a visual image, especially photojournalism that is intended to document events authentically. Having primarily AI generated visual works identified as AI is important in maintaining confidence in authentic visual imagery.

Generative AI technology also offers tremendous opportunity as it is very easy for people to create new works that can add to the richness of the visual community. In that regard, members have been developing their own generative AI models and licensing content to platforms to responsibly create models that provide revenue to creators both for the training sets and output. Importantly, these licensed datasets offer controls on what content is used for datasets as well as controlling the type of uses and permitted prompts to limit the types of output that can be created. The output of these responsible models contains metadata so that the public understands that the visuals works were created using generative AI and are not passed off as authentic.

2. *Does the increasing use or distribution of AI-generated material raise any unique issues for your sector or industry as compared to other copyright stakeholders?*

Many members have invested significantly and have built large technology platforms that enable users to search for a vast variety of visual and audio-visual content that is cleared for either editorial or commercial licensing. (Editorial licensing is primarily news, sports and entertainment content that has no releases for people or property but can be used to illustrate newsworthy, cultural and historical events. Commercial content has releases from the subjects and properties for commercial use). Because the content must be searchable for licensing purposes, the digital files contain robust metadata, including information regarding availability of releases, date of creation, locations, descriptions, captions and associated keywords. Some members have datasets that have model releases that permit biometric use, which adds additional value. The content with the associated metadata is valuable for uses in specific training models as it offers reliable information with the digital content, resulting in models that may have less bias and errors than models that are merely scraped from the internet.

DMLA members are already well-versed in licensing content on a large scale and are already licensing selected content for use in training models. Because of the value of curated content the licensing market for DMLA members datasets is already a viable market, provides revenue for rightsholders and allows for the exercise of controls on the use of the output via the licensing terms. DMLA recognizes that individual visual artists may not be in the same position and may need an association or licensing entity to license through.

3. *Please identify any papers or studies that you believe are relevant to this Notice.*

These may address, for example, the economic effects of generative AI on the creative industries or how different licensing regimes do or could operate to remunerate copyright owners and/or creators for the use of their works in training AI models. The Office requests that commenters provide a hyperlink to the identified papers.

[Not answered]

4. *Are there any statutory or regulatory approaches that have been adopted or are under consideration in other countries that relate to copyright and AI that should be considered or avoided in the United States? How important a factor is international consistency in this area across borders?*

Regulatory harmonization may be hard to accomplish, but it is key to the successful regulation of generative AI within the U.S. and on an international level. U.S. regulators and legislators must make it a priority to work with their counterparts in Europe as well as other global members of the Berne Convention to develop basic international norms and standards including basic transparency standards. As a long-standing global leader in support of the creative industries and intellectual property protection, the U.S. should stress the need to respect intellectual property and human rights when working with international counterparts. Such an approach aligns with that being taken in other jurisdictions. Even in Japan, which is one of the few jurisdictions with a text and data mining exception to copyright law, the government is taking steps to improve transparency of training data amongst AI developers. Currently, the U.S. is engaged in various ongoing AI discussions, including the G7 Hiroshima Process.

We view the most recent version of the draft EU AI Act in the European Union, as published by the European Parliament, as an effective approach to addressing transparency standards in relation to both (i) sources of training data that includes copyright works at the input stage; and (ii) the labelling of outputs of generative AI tools. We believe transparency standards will help promote responsible innovation and the EU AI Act should be closely consulted by policy makers in the US.

While other jurisdictions, including Hong Kong, South Korea, Australia, and Canada have begun to consider AI regulations and policies with respect to copyright laws, such instances are currently few and far between. These countries have considered creating copyright exceptions for text-and-data mining, but significantly, have declined to take such action, at least at present. Meanwhile, China has put forth interim generative AI guidelines, which include a requirement for generative AI services to respect intellectual property rights.

Only the EU, Japan, Singapore, and the UK have AI policies and regulations within their

copyright laws. Of this group, we would flag Singapore's as worrisome, as it overbroadly permits unauthorized mining of copyrighted works, including pirated works, for any purpose and with no allowance for an opt out by rightsholders, undermining creators' and rightsholders' ability to earn compensation for use of their works and disincentivizing them from creating further works. The UK, which once considered following this precedent, met strong opposition and walked back the idea of amending its copyright law to exempt text and data mining for any purpose. Their government has not found sufficient evidence to justify such an exception and has acknowledged that data licensing markets do exist that can facilitate machine learning at scale. As such, the UK currently only provides an exception for mining for noncommercial research.

5. *Is new legislation warranted to address copyright or related issues with generative AI? If so, what should it entail? Specific proposals and legislative text are not necessary, but the Office welcomes any proposals or text for review.*

As there is already a licensing market for DMLA members' libraries of content there should be no need to put aside longstanding copyright laws and policies to usher in rash new data mining laws or policies that obligate creators and their representatives to essentially subsidize AI technology under the guise of incentivizing the progress of the technology. This is especially so where there is no evidence of market failure or problems warranting an overhaul of existing laws and licensing models, which have applied to many industries, including media industries, throughout their existence. Affording AI systems and purveyors statutory exceptions to copyright law would effectively strip rightsholders of their ability to control their works and be compensated for their works' use throughout the process of AI ingestion.

At this time, it is not evident that new copyright legislation is needed, particularly considering the numerous cases pending in the federal court system addressing generative AI, poised to provide guidance from the judiciary. If the consensus following the progression of these cases is that the courts have misinterpreted the law, and permitted massive and systematic copying under an expanded fair use analysis, then narrowly focused legislation to Section 107 may be needed to correct any such misinterpretations.

However, a federal right of publicity law may offer uniformity and if drafted properly could be an improvement over the current patchwork of state statutory laws and common laws. Any federal bill would need to ensure that the act of licensing by a member was exempt from any definition of commercial activity that would require permission. Importantly, any federal law would need to expressly exempt the licensing of works that are protected by the First Amendment from requiring permission, specifically content that is used by customers in a newsworthy or expressive manner. The ability to draft a law that satisfies all stakeholders may, accordingly, be problematic, particularly considering the divergence among different states' laws concerning how long the right of publicity

endures after death.

Further, concerns regarding the use of generative AI to create deep fakes or “fake news” are outside the issues that current right of publicity laws are intended to redress; such laws instead focus on the commercialization of a person’s identity without consent. As such, additional legislation covering areas like AI labeling and transparency may be necessary to address the problem of generative AI outputs impersonating individuals’ likeness, voice, and speech.

Training

If your comment applies only to a specific subset of AI technologies, please make that clear.

6. *What kinds of copyright-protected training materials are used to train AI models, and how are those materials collected and curated?*

At present there do not appear to be any limits to material used to train AI; AI developers are pulling from a wide range of copyrighted works. Essentially any copyrighted work that has been digitized can be used for training. In the future, technology may evolve to permit training on non-digital materials, as well.

DMLA members are primarily partnering to create responsible generative text to image AI models. While some generative AI models are based on scraping the internet, members of DMLA have been partnering with third parties or creating their own models through licensed content that avoids many of the copyright and other ethical concerns common in those models that are based on scraping material from the internet. For example, Shutterstock has partnered with Open AI, Getty Images with Nvidia and Adobe Images has created Firefly based on licensed images. Such licensing arrangements allow for the licensor to build in responsible uses of copyrighted content, establishing guardrails that provide, for example, that the AI system cannot produce third-party intellectual property or deepfakes, and remunerate content creators for the use of their work in datasets and any subsequent use of dataset content in AI-generated output.

6.1 *How or where do developers of AI models acquire the materials or datasets that their models are trained on? To what extent is training material first collected by third-party entities (such as academic researchers or private companies)?*

As noted in the answer to Question 6 above, training materials may be acquired as proprietary works, via licensing, through illegal scraping, or from the public domain. AI companies can license or source data directly. Developers have also used prepared datasets, like LAION, Books 3, and more.

DMLA does not know the extent to which training material is first collected by third party entities; however, we recognize that issues with data laundering exist in the AI context.

6.2 *To what extent are copyrighted works licensed from copyright owners for use as training materials? To your knowledge, what licensing models are currently being offered and used?*

With every passing day, we are seeing more license agreements reached between members of DMLA and AI companies for licensing of dataset for text to image generative AI.

6.3 *To what extent is non-copyrighted material (such as public domain works) used for AI training? Alternatively, to what extent is training material created or commissioned by developers of AI models?*

(Not answered)

6.4 *Are some or all training materials retained by developers of AI models after training is complete, and for what purpose(s)? Please describe any relevant storage and retention practices.*

Retention practices vary among AI developers. Some will delete the training sets used in their AI models, while some will store them. However, retention policies do not actually have much bearing on copyright infringement. The right of reproduction may be violated regardless of whether a work is retained or stored.

However, the unauthorized storing of training materials does raise piracy concerns. It is important to enact safeguards to prevent mass piracy through training data, as pirated copies of works can be perpetuated, transferred, and put to additionally infringing uses.

7. *To the extent that it informs your views, please briefly describe your personal knowledge of the process by which AI models are trained. The Office is particularly interested in:*

7.1 *How are training materials used and/or reproduced when training an AI model? Please include your understanding of the nature and duration of any reproduction of works that occur during the training process, as well as your views on the extent to which these activities implicate the exclusive rights of copyright owners.*

(Not Answered)

7.2 *How are inferences gained from the training process stored or represented within an AI model?*

(Not Answered)

7.3 *Is it possible for an AI model to “unlearn” inferences it gained from training on a particular piece of training material? If so, is it economically feasible? In addition to retraining a model, are there other ways to “unlearn” inferences from training?*

(Not Answered)

7.4 *Absent access to the underlying dataset, is it possible to identify whether an AI model was trained on a particular piece of training material?*

The capacity to prompt an AI to produce identical output of a protected work would indicate it trained on the work. For example, one could prompt the AI to incorporate a specific visual artwork in its output. There are also tools like “Have I Been Trained” that help rightsholders discover if their work was used to train AI models. We believe that records of materials used for creating training models should be maintained by the entities that create databases for these training models.

8. *Under what circumstances would the unauthorized use of copyrighted works to train AI models constitute fair use? Please discuss any case law you believe relevant to this question.*

The systematic and wholesale ingestion of copious copyrighted material by AI systems should not be considered as categorical fair use. Determining fair use is, and has always been, a fact-specific inquiry considered on a case-by-case basis through application of the four factors set forth in section 107 of the Copyright Act. While AI companies may attempt to argue that their uses qualify as fair because they serve a “transformative” purpose—and that they may therefore avoid obtaining authorization via licensing from media aggregators and rightsholders—as the Supreme Court recently clarified in *Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith* (“Warhol”), the transformativeness of a use is not dispositive of fair use. Rather, it is merely one consideration within the first fair use factor. The fourth factor remains significant in determining the extent of the “effect upon the potential market for or value of” the works ingested under 17 U.S.C. § 107(4).

Additionally, AI companies must be required to implement safeguards to prevent infringing AI-generated outputs. Allowing prompts that call for the production of copyright-protected material or output “in the style” of are more likely to result in infringing outputs. Although merely imitating an existing artist’s style is not infringement, such prompts present the risk for the substantial similarity between a preexisting copyrighted work and generative output that could give rise to infringement. The risk of infringing outputs is another consideration in favor of AI companies licensing ingested works, because in doing so, parties can negotiate safeguards to prevent subsequent generated infringement from the outset.

Fair use is highly fact-specific and must be considered on a case-by-case basis. As such, it’s impossible to draw generalized conclusions as to blanket circumstances that would authorize use of copyrighted works, particularly considering the many ways in which an AI model is trained to use copyrighted works.

In considering the various fair use factors, however, there are certain notable trends surrounding AI. Under the first factor, which considers the purpose and character of the use, an AI

system's output will often serve the same purpose as the ingested work and may not be considered transformative, as a result. Moreover, many AI models have a commercial purpose. The second factor, the nature of the copyrighted work, would also typically weigh against fair use because the works ingested will frequently be expressive or creative, as in the case of visual artworks. The third factor, the amount or substantiality of the portion used, will often weigh against fair use because often complete and identical copies are made. The fourth factor, considering the effect of the use on the potential market for or value of the work, would consider the impact on existing and potential licenses for the copyrighted work; here AI output would fill a role as a substitute for the ingested work within the market and could cause harm to licensing markets for the original. Moreover, there are both existing and potential markets for licensing copyrighted content as AI training data, and unauthorized scraping of such content threatens these markets, often in bad faith, where content is scraped from pirated websites. Some current examples of generative AI systems that are licensing the content they use for training include: (i) Adobe with its generative AI system, Firefly, which trained on Adobe Stock images, openly licensed content, and public domain content; (ii) Shutterstock, which has a licensing agreement with OpenAI's Dall-E for training on its images, videos, music, and metadata; and (iii) BRIA AI, which has collaborated to train on licensed content with Alamy, Getty Images, and Envato.

Various existing fair use cases like *Sega* and *Google Books* offer some useful, analogous guidance for the present circumstances surrounding AI training. Such decisions, however, advised that the courts' analysis would differ if a copied work was expressive rather than, as in *Sega*, functional computer code. In *Google Books*, the court emphasized the safeguards Google used to mitigate any potential market harm, noting that its scanning and provision of books didn't usurp the market for the originals because Google's function was to provide information about the works, not substitute for the works via a competing product. With AI, market harm persists considering the active licensing markets and future potential markets for copyrighted content to be used in training. *Warhol* meanwhile recently confirmed that transformativeness is not controlling, which is essential guidance for AI companies who may otherwise have sought to rely on transformativeness to justify their unauthorized use of copyrighted material.

8.1 *In light of the Supreme Court's recent decisions in Google v. Oracle America and Andy Warhol Foundation v. Goldsmith, how should the "purpose and character" of the use of copyrighted works to train an AI model be evaluated? What is the relevant use to be analyzed? Do different stages of training, such as pre-training and fine-tuning,⁴³ raise different considerations under the first fair use factor?*

With *Warhol* the Supreme Court has rebalanced the fair use factors. Previously, certain courts had adopted a skewed balance of the fair use factors, emphasizing a finding of transformative purpose within the first as dispositive, such that a finding that a secondary work was transformative would all but lead to a finding of fair use. With *Warhol*, we now recognize that whether a use is transformative is

not controlling, either of fair use overall or the purpose and character of the use within factor one. As a result, attempts by AI developers to claim that the transformative nature of generative AI qualifies it as fair use is unsupported.

With *Warhol*, AI developers must be able to justify their use, which they cannot do where they have scraped material rather than licensed it. This is particularly so where other AI developers manage to train their AI without scraping works for use as training data indiscriminately across the Internet. *Warhol* also confirms that when a use will result in a substitute being available to the public for the original, it “undermines the goal of copyright.” The use to which a dataset is put may affect whether it is transformative, however, as with a dataset used for noncommercial purposes. The relevant use to be considered is generally the ingestion of copyrighted works by AI developers to train AI systems, considering there is already a market for licensing these works.

Google v. Oracle has less applicability to present AI considerations as its holding is limited to a distinct type of functional computer code.

8.2 *How should the analysis apply to entities that collect and distribute copyrighted material for training but may not themselves engage in the training?*

If an entity is collecting or curating a dataset and in so doing reproduces the copyrighted work without authorization, it will be liable for that act under the rights of reproduction and distribution (where it has distributed the works to others), regardless of whether it engaged in training itself. Moreover, under doctrines of secondary liability within copyright law, it may face liability even if it did not directly copy or distribute the works.

8.3 *The use of copyrighted materials in a training dataset or to train generative AI models may be done for noncommercial or research purposes.⁴⁴ How should the fair use analysis apply if AI models or datasets are later adapted for use of a commercial nature?⁴⁵ Does it make a difference if funding for these noncommercial or research uses is provided by for-profit developers of AI systems?*

It should not make a significant difference if funding for noncommercial or research uses is provided by for-profit developers of AI systems, as much research is funded by for-profit tech organizations.

However, fair use must be analyzed on a case-by-case basis, and noncommercial uses may still result in market harm and weigh against a fair use finding based on the first and fourth factors.

8.4 *What quantity of training materials do developers of generative AI models use for training? Does the volume of material used to train an AI model affect the fair use analysis? If so, how?*

The quantity of training material used generally depends on the AI model type and its type of output; however, AI systems are well-known for requiring ingestion of millions or billions of works for training purposes. Volume will not affect the fair use analysis, nor should any professed difficulties in

obtaining licenses for large volumes of work play any role, considering that brokering such licenses is already the role of media aggregators, like DMLA members.

- 8.5 *Under the fourth factor of the fair use analysis, how should the effect on the potential market for or value of a copyrighted work used to train an AI model be measured?⁴⁶ Should the inquiry be whether the outputs of the AI system incorporating the model compete with a particular copyrighted work, the body of works of the same author, or the market for that general class of works?*

AI-generated outputs may compete with and act as a substitute in the market for the ingested copyrighted works they trained on, which would harm the market under the fourth factor and weigh against a fair use finding. Under this factor, it's important to also consider not only whether a licensing market exists but whether there are potential future markets for rightsholders to exploit. We should examine whether the AI output acts as a substitute in the market for a particular work, the body of works of the same author, as well as the market for that general class of works; all are relevant. Where an AI company implements safeguards to reduce the risk that the AI will generate output that competes with an artists' entire body of work, it may be more fitting to limit such considerations to a comparison of input to output only, rather than the artists' overall oeuvre.

However, in terms of the use of copyrighted works as training data, the nature of any subsequent AI generated outputs does not bear on a claim for infringement. As stated, there already exist markets and further potential markets for licensed use of this content under the fourth factor of the fair use analysis.

9. *Should copyright owners have to affirmatively consent (opt in) to the use of their works for training materials, or should they be provided with the means to object (opt out)?*

DMLA supports an opt in approach whereby copyright owners affirmatively must consent to the use of their works for training materials, which is the default under existing law. Copyright law has never taken an opt out approach, as the law vests rights of reproduction with creators upon the creation of their work. The only way to prevent such scraping is through robot.txt, a text file that webmasters use to instruct search engine robots how to crawl and index pages on their website, but which is, however, ineffectual at truly preventing data scraping. DMLA members cannot use this fix as it would interfere with the legitimate business of licensing images and would prevent visual content from being searched for legitimate purposes.

- 9.1 *Should consent of the copyright owner be required for all uses of copyrighted works to train AI models or only commercial uses?*

Consent of the copyright owner should be—and under copyright law is—required for all uses of his or her copyrighted works for training purposes, unless a viable defense applies.

- 9.2 *If an “opt out” approach were adopted, how would that process work for a copyright owner who objected to the use of their works for training? Are there technical tools that might facilitate this process, such as a technical flag or metadata indicating that an automated service should not collect*

and store a work for AI training uses?

An “opt out” approach would prove challenging for a rightsholder to truly object to, considering that once an AI system ingests materials as data it cannot unlearn it. As such, any opt out process would not truly allow a rightsholder to remove their content from the AI system. If an AI developer ingested a copyrighted work protected by opt out measures, that ingestion should be considered willful infringement and potentially result in heightened damages under the Copyright Act.

9.3 *What legal, technical, or practical obstacles are there to establishing or using such a process? Given the volume of works used in training, is it feasible to get consent in advance from copyright owners?*

Certain existing technical means do not work sufficiently, like robot.txt. Copyright law already supports an “opt in” approach, regardless of proponents who argue that the advancement of AI technology is significant enough cause to allow AI developers to circumvent the law. However, the fact remains that AI cannot unlearn content it has ingested, making an opt out approach particularly fraught.

It is certainly possible to obtain consent from copyright owners, even on a mass scale, as this is already the practice with many licensing models and licensing entities that implement them. Moreover, rightsholders prefer licensing, as demonstrated by *Warhol*. Shying away from adhering to the law because it may theoretically be challenging to obtain consent for large volumes of works simply incentivizes infringement without taking owners’ rights into consideration.

9.4 *If an objection is not honored, what remedies should be available? Are existing remedies for infringement appropriate or should there be a separate cause of action?*

If an opt out request were not honored, it should serve as evidence of willful infringement in a statutory damages award.

Further causes of action, particularly where other legal issues like privacy are implicated, may also be warranted.

9.5 *In cases where the human creator does not own the copyright—for example, because they have assigned it or because the work was made for hire—should they have a right to object to an AI model being trained on their work? If so, how would such a system work?*

We recommend maintaining an opt-in system for all, even if a human creator does not own the copyright. The fact is that many artists and authors maintain business entities to which they assign the copyright in their works, rather than owning it themselves outright.

10. *If copyright owners’ consent is required to train generative AI models, how can or should licenses be obtained?*

Ideally, licenses should be obtained before use of copyright owners’ works. However, obtaining licenses in this context need not deviate from usual channels; interested parties may contact

owners directly or through their representatives, which is already the way various AI licenses are being obtained at present.

10.1 *Is direct voluntary licensing feasible in some or all creative sectors?*

(Not answered)

10.2 *Is a voluntary collective licensing scheme a feasible or desirable approach? Are there existing collective management organizations that are well-suited to provide those licenses, and are there legal or other impediments that would prevent those organizations from performing this role? Should Congress consider statutory or other changes, such as an antitrust exception, to facilitate negotiation of collective licenses?*

It is important that copyright owners and their representatives retain control over their works, and as such, that participation in any collective licensing regime surrounding AI should be optional, or opt-in, rather than compulsory. While individual creators not represented by a licensing entity may benefit from a solution such as collective licensing, such an approach should not be the industry-wide standard. Members who represent content on an exclusive basis should retain the ability to broker direct licensing agreements.

10.3 *Should Congress consider establishing a compulsory licensing regime?⁵⁰ If so, what should such a regime look like? What activities should the license cover, what works would be subject to the license, and would copyright owners have the ability to opt out? How should royalty rates and terms be set, allocated, reported and distributed?*

No, it is important for the operation of our free-market system that Congress ensures any collective licensing arrangement remain voluntary. Competition within the market better serves consumers and ensures superior products and services.

10.4 *Is an extended collective licensing scheme a feasible or desirable approach?*

No, any system that requires an opt-out is ill-advised.

10.5 *Should licensing regimes vary based on the type of work at issue?*

Yes, the type of work at issue, such as visual artworks, should be taken into account in the creation of licensing regimes, considering industry and work-type-specific differences that will affect the interests of copyright owners' opting into such regimes.

11. *What legal, technical or practical issues might there be with respect to obtaining appropriate licenses for training? Who, if anyone, should be responsible for securing them (for example when the*

Formatted: Normal, Indent: First line: 0 cm, No widow/orphan control, Don't adjust space between Latin and Asian text, Don't adjust space between Asian text and numbers

curator of a training dataset, the developer who trains an AI model, and the company employing that model in an AI system are different entities and may have different commercial or noncommercial roles)?

As specified, licensing within the context of AI systems does not substantively differ from any other type of licensing, particularly in terms of any legal or technical issues. DMLA members are capable of licensing on a large scale.

The entity seeking to use copyrighted content for ingestion and training purposes, often the AI developers, should be responsible for securing the appropriate licenses. Without doing so, they will be liable for infringement for any unauthorized use. However, other figures along the causal chain of AI data mining, such as dataset curators and companies employing the AI system, will also have a vested interest in ensuring authorization was granted to use any copyrighted content via licensing, considering copyright law's recognition of secondary liability and liability for downstream infringing uses.

12. *Is it possible or feasible to identify the degree to which a particular work contributes to a particular output from a generative AI system?*

Yes, it may be possible to identify the degree to which a particular work contributes to a particular output generated by an AI system, particularly if the output is substantially similar to an existing work. However, as noted, the similarities between the two works or the level to which a copyrighted work may have contributed to a particular generated output does not bear on the infringement analysis surrounding training or ingestion, as infringement arises upon the creation of an unauthorized copy of a protected work.

13. *What would be the economic impacts of a licensing requirement on the development and adoption of generative AI systems?*

The economic impacts of requiring licensing in the context of developing AI systems is not what we should be concerned with, considering that, in many instances, licensing practices are already in play and have not halted or otherwise adversely affected the development of AI technology. Particularly where, in accordance with copyright law, licensing is advanced as a cost of doing business, no one AI company may seek to outpace its competitors through incorporating infringing practices into its business model. The benefits of licensing are demonstrated across the many other industries that already abide by such legal practices and those within the AI sphere currently do.

Moreover, requiring and abiding by licensing practices in this context will actually better facilitate the development of AI systems, considering the removal of the threat of litigation, which—as we are seeing across the many current active AI infringement cases—does more to halt this technology and drain its developers' resources than licensing requirements could.

14. *Please describe any other factors you believe are relevant with respect to potential copyright liability for training AI models.*

(Not Answered)

Transparency & Recordkeeping

15. *In order to allow copyright owners to determine whether their works have been used, should developers of AI models be required to collect, retain, and disclose records regarding the materials used to train their models? Should creators of training datasets have a similar obligation?*

Yes, we support transparency around AI ingestion of copyrighted works, which is vital to ensuring that rightsholders be able to monitor the use of their works and guard against infringement, both during the ingestion, as well as output, processes. Transparency surrounding the ingestion of copyrighted works will help to ensure that rightsholders' rights are respected and will, moreover, promote ethical and unbiased AI practices.

15.1 *What level of specificity should be required?*

We advise recordkeeping surrounding the work that was copied, including the type of work and its title, the date it was copied, the use made, whether copies of the work have been retained and the duration for which they will be held, whether copies are disseminated to third parties, and any measures in place to prevent copies from being further disclosed without authorization.

15.2 *To whom should disclosures be made?*

This disclosure should be made to the rightsholder and/or licensing representative, as applicable,

15.3 *What obligations, if any, should be placed on developers of AI systems that incorporate models from third parties?*

(Not Answered)

15.4 *What would be the cost or other impact of such a recordkeeping system for developers of AI models or systems, creators, consumers, or other relevant parties?*

(Not Answered)

16. *What obligations, if any, should there be to notify copyright owners that their works have been used to train an AI model?*

For works ingested without permission, it should be the obligation of each AI company to conduct a reverse-search of its databases of copyrighted works to identify to rightsholders the inclusion of their work or works as part of its training. The onus should not be on rightsholders to undertake this effort, as they are not in the best position to understand—or compel an AI company to disclose—which one or number out of potentially billions of copyrighted works ingested were theirs. Additionally, we

recommend that AI developers make publicly available a searchable database of any URLs from webpages that were scraped and operate under an obligation to disclose whether a model was trained on a particular work.

17. *Outside of copyright law, are there existing U.S. laws that could require developers of AI models or systems to retain or disclose records about the materials they used for training?*

Some state privacy laws or laws relating to biometrics may require the retention of records.

Generative AI Outputs

If your comment applies only to a particular subset of generative AI technologies, please make that clear.

Copyrightability

18. *Under copyright law, are there circumstances when a human using a generative AI system should be considered the “author” of material produced by the system? If so, what factors are relevant to that determination? For example, is selecting what material an AI model is trained on and/or providing an iterative series of text commands or prompts sufficient to claim authorship of the resulting output?*

In line with the Copyright Office’s current guidelines, we believe there are instances where a human using generative AI should be considered the author of material it produces. Although the circumstances will be fact-specific, including in the context of creating of visual and media content, where the AI contributes minor, non-substantive adjustments, such as color correction or image sharpening, it would be appropriate to deem its human user the author of the work.

The extent to which prompt “engineering” may be deemed sufficient to afford human authorship in the use of generative AI will again depend on the circumstances. It is conceivable that a human author who inputs their own illustration or media file into an AI system as a prompt may have a greater claim to authorship, as in the case of an author who inputs a lengthy and detailed series of prompts based on their own work, refining the AI output at each step. Crucially, the substance of the prompt, as noted above, will play a major role in a determination of authorship, as where a human user directs the AI to modify color or crop or layer portions of an existing image, as opposed to generating an image from scratch. In this vein, the foreseeability of the AI’s results may bear on authorship, as where there is a limited range of specific expressive output that is objectively foreseeable as a result of a human user’s prompt. Prompts themselves generated by AI may be less likely to result in copyrightable work.

Copyrightability

19. *Are any revisions to the Copyright Act necessary to clarify the human authorship requirement or to provide additional standards to determine when content including AI-generated material is subject to copyright protection?*

It should not be necessary to revise the Copyright Act in relation to clarifying the human authorship requirement, as there has long been abundant consensus surrounding this requirement at judicial, administrative, and political levels. Nor should there be any changes required at this stage to provide additional standards regarding when content including AI-generated material is subject to copyright protection, as courts and the Copyright Office proceed to advise the public as to application of the human authorship requirement in the context of generative AI.

20. *Is legal protection for AI-generated material desirable as a policy matter? Is legal protection for AI-generated material necessary to encourage development of generative AI technologies and systems? Does existing copyright protection for computer code that operates a generative AI system provide sufficient incentives?*

Not all content, including AI-generated material, may be subject to copyright protection, nor is this protection necessary for the development of generative AI as a whole. Copyright itself is not limitless, and not all works, including unfixed works, some conceptual art, slogans, ideas, procedures, and lists of ingredients, for example, are protected by copyright. In this vein, it is not out of keeping with copyright law to limit protection for AI-generated material where the material is wholly generated by AI. Considering the wide range of applications to which generative AI may be put, such a stance does not deny copyright protection to users of AI systems across the board or disincentivize use of the technology, considering the many possible scenarios in which its use may still lead to copyrightable output.

Existing intellectual property protection, including patents and trade secrets, in addition to copyright, may serve to protect computer code and incentivize AI investment.

- 20.1 *If you believe protection is desirable, should it be a form of copyright or a separate sui generis right? If the latter, in what respects should protection for AI-generated material differ from copyright?*

(Not Answered)

21. *Does the Copyright Clause in the U.S. Constitution permit copyright protection for AI-generated material? Would such protection “promote the progress of science and useful arts”? If so, how?*

The language of the Copyright Clause notes that copyright is “secured for limited Times to Authors and Inventors,” denoting the human authorship requirement attached to copyright

protection. As such, purely AI-generated material falls outside the realm of protectability, which is all the more reasonable considering copyright's intention to incentivize authors to create work, which would not be applicable, or have any effect, upon an AI system itself.

Infringement

22. *Can AI-generated outputs implicate the exclusive rights of preexisting copyrighted works, such as the right of reproduction or the derivative work right? If so, in what circumstances?*

(Not Answered)

23. *Is the substantial similarity test adequate to address claims of infringement based on outputs from a generative AI system, or is some other standard appropriate or necessary?*

(Not Answered)

24. *How can copyright owners prove the element of copying (such as by demonstrating access to a copyrighted work) if the developer of the AI model does not maintain or make available records of what training material it used? Are existing civil discovery rules sufficient to address this situation?*

This hypothetical underscores why transparency on the part of AI developers and organizations is essential to ensuring that copyright owners may establish copying in the event of infringing output. Absent transparency requirements on the part of AI purveyors, existing discovery rules may not be sufficient, or affordable, for copyright owners to invoke in order to establish an infringement claim.

25. *If AI-generated material is found to infringe a copyrighted work, who should be directly or secondarily liable—the developer of a generative AI model, the developer of the system incorporating that model, end users of the system, or other parties?*

The owner of the generative AI model is the more appropriate figure to assign direct liability in the event of copyright infringement, with secondary infringement—vicarious or contributory liability—potentially attaching to other figures along the process, depending on the facts at issue and any indemnification policy applied to users on behalf of the model's owner.

25.1 *Do “open-source” AI models raise unique considerations with respect to infringement based on their outputs?*

(Not Answered)

26. *If a generative AI system is trained on copyrighted works containing copyright management information, how does 17 U.S.C. 1202(b) apply to the treatment of that information in outputs of the*

system?

(Not Answered)

27. *Please describe any other issues that you believe policymakers should consider with respect to potential copyright liability based on AI-generated output.* (Not Answered)

Labeling or Identification

28. *Should the law require AI-generated material to be labeled or otherwise publicly identified as being generated by AI? If so, in what context should the requirement apply and how should it work?*

Labeling of primarily AI-generated material offers both benefits and drawbacks. For instance, labeling of material as being generated exclusively by AI will assist the public, in knowing that a work is not “authentic”. For example, images take for purposes of photojournalism and used in an “editorial context” should never be generated or manipulated. If the public doesn’t know what images are “authentic” and what are synthetically created, the public may not know which images to trust and will not believe in images that are published in the news. The Content Authenticity Initiative is looking into a watermark that can be voluntarily placed on an image to disclose any manipulation and identify origin. Whether the burden to place a disclaimer on AI generated images is on the creator of those images or for non-AI generated images creator to include a watermark to establish authenticity is an issue under debate and consideration. However, overly burdensome disclosure requirements may disincentivize using AI and may be onerous if the works are not solely created using AI. However, the more information the public has regarding the source and provenance of an image, particularly editorial (in contrast to commercial) images, is beneficial in understanding what one is viewing. While creative, artistic or advertising images have always been understood to be more representational of an artist’s imagination, and may be staged, combined or otherwise manipulated with tools such PhotoShop, the integrity of editorial images is important and the public’s trust in what is being shown to them is essential to a free society.

28.1 *Who should be responsible for identifying a work as AI-generated?*

In this instance, the person who generated the output should be responsible for identifying the work as AI-generated; however, if that person is an employee using AI within the scope of their employment, the responsibility should fall to their employer. AI operators should also implement a system so that output generated using their AI models may easily be identified as being AI-generated.

28.2 *Are there technical or practical barriers to labeling or identification requirements?*

(Not Answered)

28.3 *If a notification or labeling requirement is adopted, what should be the consequences of the failure to label a particular work or the removal of a label?*

Consequences could include fines or suspension of any licenses granted; however, it would be more appropriate to first understand the contours of any labeling requirement system before proposing consequences for failing to adhere to it.

29. *What tools exist or are in development to identify AI-generated material, including by standard-setting bodies? How accurate are these tools? What are their limitations?*

The Content Authenticity Initiative (CAI) is an Adobe led initiative provides tools to identify the provenance of digital content. <https://contentauthenticity.org/>.

This initiative was started to allow content creators and publishers to offer information regarding the provenance of content, by adding a layer of information that can identify any changes. It was recognized that even if content creators could include information in metadata, metadata can be easily stripped, often by automated technology in compressing files, and the information can be changed, CAI is compliant with the technical specifications of a standards body released in 2022 by the [Coalition for Content Provenance and Authenticity](#) (C2PA) based on CAI and Project Origins, a Microsoft and BBC led initiative. Based on open- source tools, C2PA offers technical standards, providing publishers, creators, and consumers the ability to trace the origin of different types of media. Current members include Adobe, NY Times, BBC, Microsoft, and IPTC among others. Published content would display a C2PA watermark that a user could click on to view the provenance.

Additional Questions About Issues Related to Copyright

30. *What legal rights, if any, currently apply to AI-generated material that features the name or likeness, including vocal likeness, of a particular person?*

Many states have various statutory and/or common law rights of publicity laws that prohibit the use of someone's image or likeness, including voice, for commercial purposes. However, if the AI generated uses are editorial or expressive, such as in a parody or in the context of another work such as a documentary or film, these uses would be outside most current state laws.

31. *Should Congress establish a new federal right, similar to state law rights of publicity, that would apply to AI-generated material? If so, should it preempt state laws or set a ceiling or floor for state law protections? What should be the contours of such a right?*

While in theory, a uniform federal right of publicity law may be beneficial instead of the current patchwork of state laws, Congress should carefully consult with all the stakeholders of any competing interests in this area. The ability to license content lawfully must be maintained and the right

to create content for newsworthy and expressive purposes that are guaranteed under the First Amendment must be considered and balanced with the concerns of the public including actors and other public figures. If a federal law is enacted, it should preempt the various state statutory and common laws so their uniformity in licensing and use of content. NY recently amended its Civil Rights Section 50-51 to add Section 50(f) that includes a 40 year descendible right of publicity for deceased personalities that requires consent for commercial activity (expressly exempting newsworthy and expressive works and a limited deep fakes restriction on deceased performers. Importantly for DMLA members the bill expressly exempts from liability the act of licensing content under Section 50-f (d) (10) .

10. Nothing in this section shall apply to a person that offers a service that displays, offers for sale or license, sells or licenses a work of art or other visual work, or audiovisual work, to a user, provided the terms of such sale or license do not authorize such user to engage in acts that constitute a violation of this section.

32. *Are there or should there be protections against an AI system generating outputs that imitate the artistic style of a human creator (such as an AI system producing visual works “in the style of” a specific artist)? Who should be eligible for such protection? What form should it take?*

Some DMLA members have developed various proprietary generative AI platforms to facilitate the creation of responsible generative AI content. These generators may contain guardrails to reduce the creation of “in the style of” images or deep fakes and may include certain restrictions on prompts. or Training models themselves may be restricted by excluding editorial images to promote responsible AI generative outputs. Many have provided users with best practices in using the generating AI images to avoid creating works in the style of living artists or creating fake imagery of recognizable persons or places.

33. *With respect to sound recordings, how does section 114(b) of the Copyright Act relate to state law, such as state right of publicity laws? Does this issue require legislative attention in the context of generative AI?*

(Not Answered)

34. Please identify any issues not mentioned above that the Copyright Office should consider in conducting this study.

(Not Answered)

Respectfully submitted,

Nancy E. Wolff
As counsel to DMLA