



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 RIGHTSIFY

Dear Ms. Wilson and Ms. Strong:

Rightsify Group LLC (“Rightsify”) respectfully submits these reply comments in response to the U.S. Copyright Office’s Artificial Intelligence Study. The comments are presented by Alex Bestall, founder and CEO of Rightsify and Global Copyright Exchange LLC (“GCX”).

Rightsify has made significant strides in the global music licensing arena over the past decade, servicing thousands of businesses in over 180 countries, with its music receiving billions of streams on various consumer facing streaming platforms. In 2023, our commitment to innovation led to the establishment of GCX, the world’s first and largest music dataset service for machine learning research and generative AI commercial applications. Rightsify/GCX enables AI developers globally to legally clear music and launch their AI music products commercially. As the first music company to bring this solution to market, Rightsify has extensive real-world insights into Artificial Intelligence and Copyright.

The collective insights and positions of Rightsify and GCX on the matters pertaining to AI and copyright are detailed in the following sections of this document.

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 the 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?

Benefits:

- Generative AI systems can rapidly create a variety of music pieces using different genres, instruments or world styles, expanding the portfolio exponentially.
- Leveraging AI as a tool can reduce costs and increase the efficiency of composing music.
- AI can enable anyone to create personalized and custom music experiences.
- AI can generate licensing income

Risks:

- AI-generated music may infringe on existing copyrights.
- Over-reliance on AI may undervalue human creativity and decision making.
- AI generated music has the potential to flood the market with a large volume of music, potentially decreasing the overall value of copyrighted material.

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?

Unique issues for the music licensing sector are:

- Determining attribution and fair distribution of royalties for AI-generated content.
- Ensuring the quality of AI-generated content to maintain brand value.
- Addressing ethical considerations around the use of AI in content creation.

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.

(Intentionally omitted)

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?

In some countries (Israel, Japan, Singapore) we have seen reports that AI training makes copyright licensing exempt. We disagree with this approach and believe the United States should avoid this approach as this is contrary to the US copyright regime and could cause great harm to content creators. This approach from a few countries does not make legal sense as companies that create products in these countries will inevitably have infringement issues if they make their product commercially available in the United States. The rush to encourage rapid AI development through legislation could have untended consequences and damage many currently viable arts and entertainment industries.

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.

We believe the Copyright Office (and Congress) should make it clear that generative AI does require licensing for copyright just like any other media work that publicly performs or reproduces for profit. It is too soon to tell now whether the Copyright Act needs revision, but it is possible that there are some areas

that may need to be addressed (such as potential damages for refusing to removed unauthorized content from a model)

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

In our experience with music, raw audio files in a lossless format are a requirement (WAV or FLAC). Detailed metadata on each musical work is also extremely important to create high-quality models. Some key data points are genre, culture, style, mood, energy, key, tempo, bpm, instrumentation, time signature, bitrate, and chord progressions.

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)?

Developers often acquire materials from:

- Existing datasets that are publicly available on sites such as HuggingFace, Github, Kaggle, etc
- Collaborations with music companies, such as Rightsify/Global Copyright Exchange, and artists to access a rich database of music assets.
- Purchasing or licensing data from third-party vendors.

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?

- At Rightsify/Global Copyright Exchange, we offer multi-year licensing engagements for training and commercial use.
- AI Use cases that we have actively licensed are: Music Generation, Music Recommendation Systems, Music Tagging.

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?

Public Domain Works: Extensively used as they don't involve legal complexities and are free to access.
Commissioned Works: At Rightsify/Global Copyright Exchange we have done some commissioned works for AI models. However, the vast majority of content is comprised of licensed catalog works.

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 for Improvement: Developers retain training materials to further improve and refine their AI models for a set term of the agreement.

Data Warehousing: Storing data securely is a must to prevent leaks or privacy issues.

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.

During training, materials are processed and analyzed extensively. This might involve temporary reproduction of works. It could potentially implicate the exclusive rights of copyright owners, depending on the extent and nature of reproduction.

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

Inferences are usually stored as weights and biases within the neural networks of the AI model.

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 retaining a model, are there other ways to “unlearn” inferences from training?

It is technically possible to “unlearn” inferences, but it is not economically feasible for most companies. This would require retraining or fine-tuning the model which has significant cost and time investment.

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?

Without access to the raw training data, identifying if a specific piece of material was used in training can be very difficult, unless an AI generated song is generated with a very close, or exact match of an original work. This is why transparency and record keeping of the content trained on is essential.

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.

Fair use is intended to be determined on a case-by-case basis. There is no legal precedent for the massive scraping of data for the purposes of creating data sets that can be commercially exploited to potentially create competing works. This exceeds the boundaries of any current copyright cases involving the uses of databases that are used to facilitate search or other similar non-competing purposes.

8.1. In light of the Supreme Court’s recent decisions in *Google v. OracleAmerica*⁴¹ and *Andy Warhol Foundation v. Goldsmith*,⁴² 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?

As the training sets are created for the ultimate purpose of developing commercial models, the end purpose should be the only issue that matters. The intermediate steps are all part of the process and should not be analyzed separately under the first factor. As music is an expressive work, fair use should weigh against fair use under the second factor. In addition, the courts should consider the impact on the already existing and emerging licensing market that would be harmed if online scraping is widespread and continues. The use of materials to create a model is a commercial use whether it is by an institution or a for profit company.

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?

41 141 S. Ct. 1183 (2021).

42 143 S. Ct. 1258 (2023).

43 See Pre-training, Fine-tuning, and Foundation Models, GenLaw: Glossary (June 1, 2023), <https://genlaw.github.io/glossary.html> (explaining that pre-training is a relatively slow and expensive process that “results in a general-purpose or foundation model” whereas fine-tuning “adapts a pretrained model checkpoint to perform a desired task using additional data”).

There should be no difference as they could be liable under secondary liability. These companies would be facilitating the infringing activity.

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?

Since the research institutes are primarily funded by commercial entities involved in the tech industry and these researchers create the datasets in collaboration with these commercial entities for their benefit, it should not make a difference if the collection of the data and development is done by a research institute or a commercial entity.

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?

A rudimentary model could be trained on a small music dataset of as little as 1,000 songs. However, high-quality music models that can resemble professionally produced music would require datasets of at least 1 million songs to be commercially viable.

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?

The issue should be that there is a current market for licensing data for data sets for all classes of works, and the market is growing.

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)?

The "opt-in" approach would generally provide more protection to copyright owners and is consistent with current US Copyright Law. While, the "opt-out" approach is less burdensome to the aggregators for training models, it is contrary to copyright owners right to control the use of their works. A voluntary licensing model would not require an opt-out.

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

Given the close collaboration, including funding between universities, research organizations and commercial entities, especially the big tech companies, it is our view that a license (consent) should be required for all use cases, whether commercial or non-commercial.

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? 48

Although we don't support this approach, if it was undertaken a copyright owner would need adequate notice and opportunity to opt out. Using a music licensing example, any music library, distributor or organization should have an opt-out period for all copyright owners they work with prior to delivering any content for an AI licensing deal. The licensee (AI company) should also be provided evidence of the opt-outs.

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?

We believe this is feasible. An automated email to all copyright holders of a party licensing works for AI training should be able to get opt-out confirmations relatively easily.

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?

Because it is difficult for a model to unlearn, it would not be reasonable to just have the content not used going forward. The damage remedy should be sufficient enough to deter flagrant ignoring of opt out because damages are too low or too difficult to prove. Perhaps statutory type damages and attorney's fees

should be considered or mandatory small claims court as a requirement to take advantage of opt-out. This is an example why an opt-out is not recommended.

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?

Because a work is created under a “Work for hire” arrangement does not mean a human has not created the work and should not object to inclusion in the model. Many creators operate under a “wholly-owned entity” or similar entity for many various business reasons but it does not mean there is no human author who benefits from the exploitation of the work. Works created under a work for hire regime should not be excluded from objecting to any opt-out.

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

It is our view at Rightsify/Global Copyright Exchange that direct licensing is the most effective way of licensing for AI training. It is our concern that traditional collective management organizations would not be able to accurately account and pay royalties for AI licenses. This is a new licensing model that requires high quality metadata for every musical work which is something collective management organizations have struggled with.

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

For music, we believe that direct licensing is feasible and the better option. However, we can’t speak for all creative sectors.

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?

We believe there is no collective management organization that is well-suited to provide licensing for the music industry. The reasons being:

- Collective management organizations administering agreements with AI companies would prevent copyright owners from effectively auditing AI models, building relationships with AI developers and would give preferential treatment to major labels and publishers.

10.3. Should Congress consider establishing a compulsory licensing regime? 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?

We believe that a compulsory licensing regime for AI music would be an ineffective way of licensing musical works and would devalue musical copyrights.

10.4. Is an extended collective licensing scheme a feasible or desirable Approach?

We do not believe so. This would likely lead to many copyright owners unwillingly granting the rights to AI licensing and the accuracy of which works were used in models and how royalties are calculated would be in question.

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

Yes, the type of content will inform the licensing regime. For music, the use cases are varied from music generation to remixing, synchronization in video games and fitness products. There is not a one-size fits all approach to AI licensing which is why we believe the collective management model is not feasible for AI.

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 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)?

The company building and/or releasing an AI model should be responsible for obtaining the necessary licenses. Per our answer for 9.1, the lines between commercial and non-commercial AI research and development are very gray, so whether it is an independent researcher or a large company, the individual or entity responsible should be obtaining the necessary licenses.

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? Please explain.

There are ways to track the similarity of an original work but there is no specific way to detect how much an original work contributes to an AI generated output.

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

A licensing requirement would increase costs for developing AI generative systems. However, it would significantly reduce legal and ethical risk for AI development and would lead to new revenue streams and opportunities for copyright owners. It may also result in better models with less bias and incorrect responses as the metadata would be more accurate.

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

Determining the threshold of originality in AI-generated music, and devising a framework to address potential infringement issues efficiently. Since fair use is a balancing of 4 factors so I don't think any other guidelines can be developed under the statute].

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?

It is our view that developers and creators of training datasets should be required to maintain detailed records of the materials and metadata used for training to ensure transparency and compliance with copyright laws. Model auditing is vital for AI licensing.

15.1. What level of specificity should be required?

A high level of specificity should be required to allow for accurate tracking and compensation of copyright owners. This includes Track Titles, Artist Names, ISRC Codes, and other unique identifiers of each musical work.

15.2. To whom should disclosures be made?

Disclosures should be made directly to the copyright owners or their designated representative.

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

Developers using models from third parties should do thorough due-diligence of any third-party models to ensure copyright compliance and should bear responsibility for any infringement. Make sure that the models are based on lawful collection and transparency about what works are used in the models?

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?

Implementing such a recordkeeping system would increase costs for developers but it would also foster a more transparent and lawful system. The developers are receiving substantial funding that should easily cover these costs.

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

If a copyright owner has willingly licensed their works and has not opted-out there should be no need for a notification of a work being included in an AI model, other than to make reporting and payment for the works that are licensed, unless otherwise agreed in the licensing contract.

If the works were ingested without permission, it should not be the burden of the copyright owner to determine if the work has been ingested and the AI model should respond to any requests for information.

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?

There may be data protection and privacy laws that could potentially have an impact on record retention and disclosure of AI models.

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?

Human involvement in the creation process, such as providing inputs or fine-tuning the outputs, could potentially be a factor in considering them as authors.

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?

No change is necessary to clarify the human authorship requirement (especially in view of the *Thaler* decision)

No change to copyright law is required at this stage to provide additional standards for determining when content from an AI system is subject to copyright protection.

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?

Legal protection might be necessary to encourage innovation and investment in the AI sector. It could be a form of copyright or a separate sui generis right with specific provisions tailored for AI-generated content.

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 sure if protection is warranted. If so, there should be a study on what it should cover and for how long. If protection is granted, it should consider the unique nature of AI-generated materials.

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”?⁵² If so, how? No reading of the clause would support non-human authorship.

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?

Yes, AI-generated outputs can potentially infringe upon the exclusive rights of pre-existing copyrighted works, especially in cases where substantial portions of such works are replicated.

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?

The substantial similarity test might still be relevant, but additional standards might be necessary to address the complex issues surrounding AI-generated content.

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?

Proving copying might be challenging, especially if records of training materials are not maintained. Current civil discovery rules might be inadequate to address this issue fully. This is why transparency is critical. There may be a need to have legislation regarding transparency and record keeping.

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?

Like all infringement questions, the answer to this question is fact-dependent.

Depending on the facts, any of the parties identified in the question can be liable under any of the existing liability principles in copyright law of direct, vicarious, and contributory liability depending on the facts at issue.

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

Open-source models could potentially lead to widespread infringement issues due to the decentralized nature of their usage.

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?

- 1202(b) applies to both the ingestion of copyrighted materials that may be stripped of CMI and the generation and distribution of potentially infringing output that contains altered or false CMI (or from which CMI has been removed).

27. Please describe any other issues that you believe policymakers should consider with respect to potential copyright liability based on AI-generated output. Labeling or Identification

Other issues might include establishing clear guidelines for determining infringement, developing mechanisms for compensation of copyright owners, and fostering collaboration between stakeholders to address copyright issues effectively.

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?

A labeling requirement might be necessary to ensure transparency and to assist in identifying potential copyright infringements. However, we are not sure how feasible this is in reality. There needs to be a distinction between works that are generated almost entirely from AI without human authorship and Works in which creative content is created using AI as a tool with additional human creativity. It could be very burdensome to label any work that has used generative AI in creating part of the components.

28.1. Who should be responsible for identifying a work as AI-generated? Some AI models are released by their developers for download and use by members of the general public. Such so-called “open-source” models may restrict how those models can be used through the terms of a licensing agreement

The entity releasing the AI-generated content should be responsible for labeling it appropriately.

28.2. Are there technical or practical barriers to labeling or identification Requirements?

There might be technical challenges in implementing a labeling system, especially in determining the threshold for labeling. Over labeling could be very burdensome.

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?

Failure to label or removal of a label could potentially result in penalties to encourage compliance.

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?

Various tools are being developed for this purpose, but their accuracy and limitations would need to be assessed continuously as technology evolves.

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?

Current laws protecting personality rights might potentially apply, but there might be a need for specific provisions addressing AI-generated content to prevent fraud.

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?

Considering a new federal right might be necessary to address the unique issues surrounding AI-generated content, with provisions to harmonize it with existing state laws.

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?

There might be a need for protections against imitation of artistic styles to protect the rights and interests of human creators.

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?⁵⁴ Does this issue require legislative attention in the context of generative AI?

Section 114(b) might have implications for state rights of publicity laws, and legislative attention might be necessary to address potential conflicts and gaps in the law.

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

