

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 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?

Potential benefit: Allowing access to creations or ideas without having to wait for the copyright to be lifted, with the potential of sparking further derivative works. Supporting creators skilled in one area (writing) advance their craft in another area with little technical know-how (story-based art, comics, animations, etc.)

Potential risk: Counterfeit works created “in the style of...” that could be passed off as original works of a creator but the profit and claim of which does not belong to the creator. Numerous uses of this would make creators wary of posting their creation on a public forum.

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?

AI-generated material? No, as work created from it is considered public domain and free for use. AI-generated material based on copyrighted work? Yes, because the possibility of unknown copyright infringement presents many legal concerns.

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.

Yes, the use of copyrighted material without the explicit permission of the creator, should not be allowed.

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?

Only copyright-protected material that have received explicit permission from the creator should be allowed for use as training materials. The “Fair Use in the Classroom” clause does not apply to this situation, as use of copyright-protected material stands the risk of devaluation for this use case [1]. This is because material used for training can then be accessed by a query of “Harry Potter, book 1, by JK Rowling”. This free access of copyrighted material is a direct act of copyright infringement as the work has been “reproduced, distributed, performed, publicly displayed, or made into a derivative work without the permission of the copyright owner.” [2]

If a creator has given written permission for their works to be added as training materials with this potential explained to them in advance, then they have waived the right to file for copyright infringement in this specific case.

[1] <https://welch.jhmi.edu/get-help/can-i-use-images-i-find-my-presentationclass-without-having-get-permission#:~:text=Effect%20of%20the%20use%20on%20the%20potential%20market%20for%20or%20value%20of%20the%20copyrighted%20work.>

[2] <https://www.copyright.gov/help/faq-definitions.html#:~:text=What%20is%20copyright%20infringement%3F>

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)?

All works currently in the public domain, of which there are many and varied examples [1], would be available for training materials. Public datasets and forums are also available for use [2] as well as specific language models such as Project Gutenberg [3] which contains an extensive collection of book texts that are within the public domain.

As such, there are many publicly available sources for training materials that do not require copyright-protected materials.

[1] <https://web.law.duke.edu/cspd/publicdomainday/2023/>

[2] https://hadoopilluminated.com/hadoop_illuminated/Public_Bigdata_Sets.html

[3] <https://odsc.medium.com/20-open-datasets-for-natural-language-processing-538bfaf8e38>

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?

As far as I am aware and am able to research, there is no current model for licensing copyright works for use as training materials. This is because licensing most often requires a limited number of uses.

For example, MATLAB is a mathematics and graphing software offered by MathWorks [1]. For a single person at a company to use the MATLAB software, the company must pay for a yearly license. This license is for the sole use of that single person. This license cannot be split across two people or a group of people. This license is a one-to-one ratio.

In a situation where copyright-protected material is licensed for training material for AI, there is no control on the number of people able to access this material. The AI software is used by as many people as a server is able to handle, which can be 500 visitors or more in a single second [2]. If only 1% were to access the licensed materials, that would still be a one-to-many relationship that makes for inconsistent rates of use. This kind of model would not be sustainable for long.

[1] <https://www.mathworks.com/company/newsletters/articles/the-origins-of-matlab.html>

[2] <https://cloudkul.com/blog/what-is-concurrent-users/#:~:text=CPU%20Cores,website%20at%20the%20same%20time.>

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.

AI stores all data it comes into contact with as a means of continual training and improvement. To get rid of any data, an AI must be re-trained from scratch [1]. This became clear when a new Art-protection program was created to make 'invisible' watermarks on copyright-protected art [2]. These 'invisible' watermarks slowly poisoned the data within an AI's program until the results were incomparable to the desired output. In order to recover, AI models would need to be retrained in an extensive and time-consuming manner because enough 'good' data must be added to outweigh the 'bad' [3].

As such, no data is ever truly deleted [4].

[1] <https://fortune.com/2023/08/30/researchers-impossible-remove-private-user-data-delete-trained-ai-models/>

[2] <https://cs.uchicago.edu/news/uchicago-scientists-develop-new-tool-to-protect-artists-from-ai-mimicry/>

[3] <https://www.csoononline.com/article/570555/how-data-poisoning-attacks-corrupt-machine-learning-models.html>

[4] https://proceedings.neurips.cc/paper_files/paper/2019/file/cb79f8fa58b91d3af6c9c991f63962d3-Paper.pdf

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?

No, see 6.4 and [1].

[1] https://proceedings.neurips.cc/paper_files/paper/2019/file/cb79f8fa58b91d3af6c9c991f63962d3-Paper.pdf

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?

Potentially. There is an AI called “MARLOWE” which advertises the ability to compare created works against previously compiled data [1]. While it may seem counter-intuitive to use an AI to analyze an AI, this shows that there is a way to track certain identifiers in created works and draw close comparisons between an AI-generated work and a copyright-protected work. In the sample of the Pro Report generated by MARLOWE [2], pages 30-32 show a break-down of comparison between the input given to MARLOWE and previously accumulated works, as well as what factors were used in this comparison. Works made by the same author appear to have higher comparison factors than works created by a different author.

[1] <https://authors.ai/marlowe/>

[2] <https://authors.ai/marlowe-pro/>

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.

See 6.

8.1. In light of the Supreme Court's recent decisions in *Google v. Oracle America* ⁽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?

See 6.

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 entities collect and distribute copyright-protected material without the permission of the author, then they are implicit in the act of copyright infringement, regardless of whether the material is used for training or not.

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?

As data cannot be deleted from AI models (see 7.3), any models or datasets that are made for research purposes cannot be later adapted for commercial use. Any model that has an eventual goal of being used for commercial use should be treated as such.

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 materials depends on the degree of accuracy and content of desired result. 100 images are enough to train an AI to recognize basic human facial features, but far more would be needed to differentiate between a stock of body care products at a grocery store [1].

The volume of the data used to train an AI model does not affect the fair use analysis. As previously stated, the AI violates the fair use clause since it allows the free reproduction and distribution of copyright-protected works without the permission of the creator. (See 6.)

[1] <https://blog.theos.ai/articles/how-many-images-do-i-need-for-my-computer-vision-model#:~:text=Generally%2C%20the%20more%20data%2C%20the,5k%2C%2015k%20and%20so%20on.>

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?

See 6.

In regards to 'derivative works', it is possible to input a query of "Harry Potter, Book 1, in the style of Stephanie Meyer". In this situation, the result would be a book with all of the characters, places, and events of Harry Potter and the Philosopher's Stone (Sorcerer's Stone for US publications) in the 'voice' of Stephanie Meyer, the author of Twilight. If Stephanie Meyer were to do the same, writing a version of Harry Potter in her 'voice' complete with characters, places, and events, JK Rowling would be able to sue for copyright infringement.

As such, fair use does not protect the use of copyright-protected material to train an AI if a similar, human-use situation would be considered copyright infringement.

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)?

Copyright owners should have to affirmatively consent to the use of their works for training materials as AI models cannot delete works once they have been added. Allowing AI developers to add material without permission from copyright owners will create more work for AI developers than requesting permission in the first place. This is not an instance of "ask forgiveness rather than permission".

9.1. Should consent of the copyright owner be required for all uses of copyrighted works to train AI models or only commercial uses? ⁽⁴⁷⁾

As a software engineer, I have seen many educational and research models end up released as 'open-source' material to be used for commercial use later. Many times, all that is requested is a note of origin for the open-source matter that is used. As such, I believe that all AI models should be viewed with the idea they will be eventually released for commercial use and so should require the consent of copyright owners in advance.

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⁾

For an “opt out” approach, a flag or metadata field would need to be added to every avenue that a copyright material appears in full. This would be every eBook website, every training data ‘scraper’ to every be created and any PDF-to-Digital converters that might be used.

Right now, there are no technical tools that have access to all corners of the internet as every website has its own administrative domain that is not available to the public. The closest similar service would be ones that erase personal information from data brokers and mass marketers [1], but these are services that are retroactive and constantly monitor the internet for ‘added’ information.

In addition, the original work *and any derivative work already generated by the AI*, must be deleted from the AI model. As stated in 7.3, deletion of material from an AI model is impossible, so the entire AI would need to be deleted and started over from scratch.

Furthermore, this would need to occur for each and every work the author has created. In the case of popular writers such as Stephen King, who has written 65 books thus far, this kind of process would have far-reaching consequences.

[1] <https://proprivacy.com/privacy-service/guides/how-to-remove-yourself-from-the-internet#:~:text=Best%20services%20to%20remove%20yourself,search%20sites%2C%20and%20mass%20marketers.>

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?

There is currently no tool available for an “opt-out” process and as deleting a copyright-protected work from an AI training model would destroy the model, the process itself would be undesirable to developers that have spent the better part of weeks to years working on their model.

If a common site were to be made available for creators to “opt-in”, I believe it to be feasible to receive consent in advance from copyright owners.

In the case of MARLOWE, this is an option given to Pro members. Any Basic members are warned prior to application that their work will be retained for training materials. Pro members have the option to opt-out their work from this retention [1].

[1] <https://authors.ai/single-report/> (An account/sign in may be required to access this page)

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?

I believe the current remedies for copyright infringement (impounding and disposition of infringing articles, damages, profits and statutory damage, along with attorney fees [1, pg. 9]) are appropriate. However, determining the amount of damage or profit could be difficult if the number-of-accesses to the copyright-protected material are not tracked and the AI model containing the infringement was used by multiple parties “unknowingly”. In such a case, while the original creator would have suffered damages, there might be no liable party for restitution.

[1] <https://www.copyright.gov/title17/chapter5.pdf>

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?

If the creator has made a work for hire, then it is within the owning company’s right to use the content as desired. If the owning company does not want it to be used as training material, the company should be allowed to follow a similar process as a copyright owner.

As this is a contract between the creator and the hiring company, there should already be similar guidelines in place.

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

See 9.3.

However, the term 'licenses' should not be used as the common terminology implies an equal exchange of monetary value in return for access to IP. As AI models can be accessed by hundreds of users at once, there is no equal exchange of 'loss' accrued due to leasing of copyright-protected material.

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

See 9.3 and 10.

I believe a direct, voluntary agreement can be accumulated from creators. It may also be possible to provide a flag or metadata field of "AI Training Welcome" to allow for more automated collection of training materials. However, a 'licensing' agreement for monetary gain is not feasible for reasons stated in 10.

I do not believe it is feasible for an entire creative sector to voluntarily agree to application of copyright-protected materials being used for training AI models. After all, everyone is entitled to their opinions and should be allowed to refuse even when their sector opts-in.

10.2. Is a voluntary collective licensing scheme a feasible or desirable approach? ⁽⁴⁹⁾ 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?

I do not believe a collective licensing scheme is feasible nor desirable for reasons stated in 10.

The closest applicable organization would likely be a Guild or Union of some kind. However, Unions only negotiate between workers and owners and Guilds negotiate between their own members [1]. In this situation, the AI developers would not be required to be a part of a Guild, nor would they be employing creators like a Union would dictate.

A new kind of collective organization would need to be facilitated in this situation and the powers of such an organization fleshed out.

[1] <https://guide.unitworkers.com/whats-a-guild/#:~:text=Unlike%20unions%20that%20focus%20on%20bargaining%20between%20workers%20and%20owners%2C%20guilds%20primarily%20facilitate%20negotiations%20between%20their%20own%20members>

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?

Congress should not be able to establish a compulsory licensing regime, as that implies that Congress owns (in part or in whole) the copyright-protected material in question. In addition, unless the AI model keeps track of the number of times a copyright-protected work is accessed for use in a query, there is no way to calculate a royalty rate for compulsively added works.

Copyright owners should always have the ability to opt out, especially if they think the use of their copyright-protected material is in anyway harmful to their profits or brand. Since the removal of

material from an AI's training database is not easy (see 7.3), this act of opt-out would be harmful to AI developers as well which might prompt legal action in recourse.

10.4. Is an extended collective licensing scheme ⁽⁵¹⁾ a feasible or desirable approach?

I do not believe a collective licensing scheme is feasible or desirable for reasons stated in 9.2, 10, and 10.3.

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

I do not believe a licensing regime would be feasible or desirable.

If one were to be made, however, there would need to be a case-by-case evaluation. The book *The Very Hungry Caterpillar* by Eric Carle would have a very different rate compared to a book such as *The Mysteries of Paris* by Eugene Sue [1]. Both books should not be compared to paintings by Monet or the once-copyrighted performance of *The Nutcracker*.

[1] <https://irisreading.com/8-very-long-books-worth-the-time-theyll-take-to-read/>

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)?

In the above options, I believe the curator of the training dataset holds the responsibility of obtaining appropriate licenses. If a dataset is made available and open-source, I (as a developer) would assume that all data within the dataset is free for use. Following that line of thought, if I (as a developer) were to make my AI model open-source, then a company that decided to use it would also assume that any data used in training were also open-source. This follows the associative property of mathematics [1].

Likewise, if a curator asked a copyright owner for permission to add their material to a dataset, with the expressed desire for AI training, then the copyright owner has been properly warned about the uses of the dataset. At this point, the copyright owner may choose to accept or reject the request.

If a curator does not obtain permission before adding a copyright-protected work to their dataset, then they are falsely advertising the dataset. It is not a free and open-source option nor, in the case of commercial use, is the dataset appropriately licensing the data it is providing.

However, it should also be noted that there is a user-element as well. At any time, if the AI is made available to non-developers, a user could enter copyright-protected material they do not own. As an example, a user who is not JK Rowling could enter the contents of Harry Potter Books 1 through 7 to see what an 8th book would generate. In such a case, there should be measures in place, such as query word-limits or flags to denote an excessively large input. These flags can later be reviewed for copyright infringement claims.

[1] [https://www.mometrix.com/academy/associative-property/#:~:text=The%20associative%20property%20states%20that%20when%20adding%20or%20multiplying%2C%20the,%C3%97\(b%C3%97c\).](https://www.mometrix.com/academy/associative-property/#:~:text=The%20associative%20property%20states%20that%20when%20adding%20or%20multiplying%2C%20the,%C3%97(b%C3%97c).)

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.

Yes, to a degree. See 7.4.

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

The economic impacts of a licensing requirement would likely slow down adoption of generative AI system as many free-lance developers would not be able to afford the licenses required for a large system like AI models. AI models require a lot of hardware and maintenance for continued use and improvement. There are already talks about monthly fees for a professional version of ChatGPT [1]. If every AI model had a similar licensing requirement, the smaller businesses would likely not adopt them until prices decreased, leaving much of the heavy development work to the large corporations capable of investment.

[1] <https://www.pcworld.com/article/1481398/chatgpt-may-charge-42-mo-for-a-paid-tier-as-microsoft-invests-again.html>

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?

I believe that the developers of AI models should collect, retain, and disclose records of which training datasets they use and if they added any data on top of the data sets. Records of materials within a dataset should be collected, retained, and disclosed by the creators of the training data sets.

As much of this information can be accumulated automatically using easy-to-develop scripts, I do not think this is a burden on the developers and curators.

15.1. What level of specificity should be required?

This field may change as situations arise, but some fields that come to mind are:

Article added, date added, article's original creator, article's original creation date, number of accesses to article, a relation or link to the results of the accesses to the article

15.2. To whom should disclosures be made?

Any copyright owner with regards to their own work, any lawyers in regards to a potential copyright case.

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?

The initial cost would likely be similar to other forms of data entry. Once inside the dataset, the impact should be minimal. As much of this information can be accumulated automatically using easy-to-develop scripts, I do not think this is a burden on the developers and curators.

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

See 11.

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?

I believe there is at least one instance where a human using a generative AI system could be considered the “author” of the material produced by the system. In a situation where *all* of the training material has come from the same source and that source prompts the AI for a result, I believe that source could claim ownership of the results.

In other words, if I trained an offline AI model using only my written works (in the forms of books, reports, letters, play scripts, etc.) and then prompted the AI to design an output based on a prompt, then I believe the work could be considered “mine”. This would be because the AI only has access to my works. As soon as a 2nd source were provided, I could no longer claim fully rights to the result of the AI model.

In the event of a 2nd model, a “co-author” might be claimed, so long as the 2nd source provided consent to having their work added to the model and had a hand in crafting the Query used to prompt the AI’s output (such as “xyz situation with every odd chapter being written by author 1 and every even chapter written by author 2”). As both the source material and the Query used for creation of the result were split evenly, I believe a “co-authorship” would still be feasible.

The same could be said for 3 creators, 4 creators, and so on, as long as a relatively even amount of source material were used in training and equal amount of effort were placed into the Query.

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?

Potentially, in accordance with the situation explained in 18.

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?

If AI-generated material comes from an AI model trained on material from the Public Domain, then the results would fall within the Public Domain. If AI-generated material comes from an AI model trained on material from a sole-source or limited sources, with full permission, then that material should be copyrightable by that sole source or limited sources (see situation described in 18).

I do not believe the development of generative AI is dependent on the ability to copyright the generated material. There is an entire culture of fan-fiction that is unable to be copyrighted, but is still thriving today. In a similar fashion, AI developers do not create AI for the ability to copyright the results, but to be able to push limits and create things they would not be able to, whether by inability or lack of skill. As such, I do not believe legal protection of results is necessary to encourage development of generative AI technologies and systems.

I believe current copyright protection of computer code provides enough incentive for Generative AI systems. As long as there is a way to copyright the code and create IP, then companies will continue to push for development of generative AI systems.

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?

See 18, 20.

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?

I do not think copyright protection for AI-generated material would either promote or degrade the progress of science and useful arts. The use of copyright-protected material would, as that would drive more creators to keep their art private to keep from unpermitted use of their work.

In the situation described in 18, not being allowed to copyright a sole-source or limited-source result would mean that the work would never be published, but without knowing the contents of the work, it is hard to claim a 'degradation' of useful arts.

In terms of science, see 20. AI-generated material does not need to be copyrighted to be appreciated and, if used by companies for internal analytics or research, it is likely such reports would fall under company property or trade secrets, outside the realm of copyright.

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?

Yes, see 7.4 and 8.5.

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?

More definitive tests would need to be run with a larger dataset, but I believe there is potential in being an adequate test to address claims of infringement.

If the curator of a training set also kept records of number-of-access and relation-to-query, this would also be clear evidence of outputs infringing on training data. However, this would imply the output is known to be from a generative AI system and which model it is from.

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?

See 22.

This is a good question, as curators and developers cannot always track what users add to the Query. In some cases, a user might add a copyrighted work they do not own to a Query to see what happens. For instance, a user who is not JK Rowling may enter the contents of Harry Potter Books 1 through 7 to see what an 8th book would generate. In such a case, there should be measures in place, such as query word-limits or flags to denote an excessively large input. These flags can later be reviewed for copyright infringement claims.

In this situation, it may also be as simple as asking the AI "Is there a previous Query that contained a 95% match or higher to..." and seeing what results are triggered. As no data is deleted from AI datasets, the AI would likely find the Query much faster than a developer looking through the logs. This kind of Query should be done with an off-line model, to be destroyed afterwards to prevent retention of the Query containing the data-to-match.

If a developer refuses to make available the records or training materials after a clear warrant has been obtained, they would be obstructing the collection of evidence. Tampering with records and training material would also be considered tampering with evidence.

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?

This would depend on where the copyright work was introduced.

If it was introduced in the training dataset, then the curator is responsible.

If it was introduced in the model by a developer, then the developer/company is responsible.

If it was introduced by an end user of the system, then that end user would be responsible. This case may make end users caution, but as long as the appropriate warnings have been placed in the End User's License Agreement, then they have been forewarned. Tracking down end users would be difficult if the AI software does not require an account linked to a usable email address.

25.1. Do "open-source" AI models raise unique considerations with respect to infringement based on their outputs? ⁽⁵³⁾

"Open-source" AI models will still be developed the same as "Closed-source" models. If the training dataset is public domain, then the AI models' output is public domain. If the training dataset is sole-source or limited-source, then the AI models' output is sole-source or limited-source. The code being "open-source" does not change how the inputs affect the outputs. The only aspect that would change is that there is no larger backing company to support development, only individuals that desire to develop free tools for others to use.

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?

If an AI has been trained to report with a copyright management system in mind, output should include a tag or information line similar to a bibliography where all copyright information used in a response is accumulated.

As users can request for this feature to 'turn off' on output, the AI cannot be held liable for ceasing function of this feature, nor the developers held accountable for the omission of copyright information in outputs.

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?

Yes, to properly capture origins of creation.

A caption containing the AI system used, the version, access date, and query used to generate the result would be the most contextually accurate. For example, "Dall E 2, v. 2.x.x, Accessed Sept. 22, Generated from "A Labrador with a frisbee"".

This way, if the picture is re-used elsewhere, the original point of creation can still be tracked.

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

Initial work should be identified by end users that created the query to generate the result. Developers may help in this task by providing automatic 'captions' for the users to copy, based on the information available.

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

I do not believe so, especially if developers help by making information such as AI name and version available in an auto-generated caption.

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?

If no caption is provided, the consequence falls to interested parties that believe a work might be an AI-generated mimicry. Confirmation should come from records of materials and queries kept by developers. If records are not found, AI confirmation, from either the source or a 2nd validation-AI might be required.

As AI becomes more prevalent, it is not feasible for every image, sound bite, or created work to be scanned for AI-generation. There will likely be false-positives, if a specific creator has been targeted more than once. So, the first alert should come from interested parties and investigated from there.

If it is found that the work is AI-generated and presents copyright infringement, then further steps should be taken.

If the work is AI-generated, but does not present copyright infringement, then there are no further steps.

If the work is not AI-generated, then the typical steps should be taken.

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?

I currently only know of MARLOWE specifically, but a quick Google search shows several other options:

[1] <https://copyleaks.com/ai-content-detector>

[2] <https://zapier.com/blog/ai-content-detector/>

[3] <https://www.searchenginejournal.com/google-unveils-new-tool-to-detect-ai-generated-images/495082/>

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?

I believe this should follow the “Street Photography” [1] rule in most situations, with the exception of AI-generated material that result in defamation or degradation of a brand.

As AI becomes more developed, there is the potential of creating realistic “photo-shopped” images that place persons of interest or celebrities in compromising situations that never occurred. In this instance, images should be checked for AI-generation earmarks, then the correct course of action taken.

[1] <https://expertphotography.com/street-photography-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?

In regards to the public, I do not know how a basic federal right would compare or conflict with the “Street Photography” rule. While this would allow for more privacy on average, not every citizen will have the tools to maintain a good-standing while also doing their jobs. Journalists and photographers in general may have many issues maintaining a livelihood if such a law were to be enacted.

In regards to Actors and Performers under temporary contract to a Studio, their likeness *should not* be procured under false pretenses. They should be able to opt-out of requirements of allowing their likeness to be copied and retained after their contract ends. Studios should not be able to retain the likeness of Actors and Performers to be used without their permission after the contract ends.

I believe it should at least set a minimum for state laws. This is because the use of AI models will span state- and country-borders by its base design. A photo taken in a ‘free’ state may appear in a ‘regulated’ state, which means that the action of taking the photo has crossed the border. Caution should be used with tools that do not have a physical limitation.

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?

Yes, there should be protections against AI systems generating outputs that imitate the artistic style of a *living* human creator. This protection should be available to anyone attempting to make a living off their work and who have not ‘opted-in’ to have their work added to a training database. This protection does not extend to work done for-hire by a company, as that work is then owned by the company.

This protection should follow similar guidelines as explained in sections 9, 10, 18 and 20.

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?

I believe it does, to an extent.

If a creator were to make a very specific sound bite that was able to be copyrighted, then it should be within their right to refuse to add their work to a training database. This type of sound bite can include: an original song, their voice whether singing or speaking, or other derivative work.

As these sounds can be used by generative AI systems to create a situation where their ‘voice’ is heard saying something they never said, this situation could lead to legal action in cases of defamation or false testimony.

If a sound is easy to reproduce, such as a balloon popping or paper tearing, then sounds could be curated for-hire by a company, which would circumvent any issues of copyright. Sound creators could also choose to add their sounds to a training dataset for free, as several sites already provide service for [1,2].

[1] <https://freesound.org/>

[2] <https://pixabay.com/sound-effects/>