AI Comments for USCO

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?***

As a TV writer, GAI systems have the potential to completely wipe out the market for my creativity. If companies are able to gain the ability to type into a machine and within seconds receive a work that is as good or better than what I or my peers could do, there will cease to be any market for human writers, overnight. These systems will rob myself and authors around the world of the very ability to feed our families. We are already seeing GAI erode jobs in copywriting, which is how I supported myself before I became a professional TV Writer.

The reality is that we should not be able to stop the evolution of technology just because someone invents a better way to do our job, as painful as that process might be. But I believe that this is situation is more complex. Cars are a more efficient mode of transportation than horses, and thus they replaced horses. But cars don't require the blood of horses to run, which is what GAI does when it gobbles up the copyrighted work of thousands of writers, artists, scientists and professionals all over the world, without compensation. And without data, AI algorithms are useless. Generative AI runs on stolen work. Full stop.

But I think the impact will be far more profound than writers having to find a new way to support themselves. In the most recent writer's strike, California lost an estimated 5 billion dollars. Currently, creating narrative video requires an army of professionals -- from camera operators, to casting directors, to the restaurants and caterers. If this technology spreads unabated, it will begin to spit out fully formed video and will devastate the market for filmed entertainment. I don't mean to be alarmist, but in that scenario, Los Angeles as a city will likely collapse. And San Francisco, a city built of the use of computer code, which these systems will also become experts in, will likely follow. Yes, there will be companies that profit from the technology, but the human cost of it could be very well be at a massive scale.

The potential downsides go far beyond economics. GAIs use a form of statistics to predict what a human might write, or draw or film. They are human impersonators. The dangers of such a tool are incredibly wide ranging. The internet was created as place where humans can connect and share information. A place to build a truly global community. But as these systems proliferate, there will be no way to know whether anything you read or see was generated by a human being. GAI is currently being used to impersonate people's loved ones in an attempt to scam them out of their money. They will be used to spread political disinformation, raising massive implications for integrity of our democracy. AI has also been used to place woman's faces on pornographic images, creating a new class of revenge porn. The internet as we know it could become a place by and for synthetic computer systems, not human beings.

Now, of course, it is natural for someone who has their livelihood threatened by a technology to not like it. But the reality is that a rather large proportion of the AI development community itself is scared of what they are building. 1,100 leaders within the technology industry recently released an open letter urging a moratorium on AI development for at least *six months* because they believe the danger is that grave. Geoffrey Hinton, whose work laid the foundation for modern AI, recently admitted in a podcast that the truly best course of action for humanity in regards to AI would be to stop developing it entirely. The very people who have built and are building this technology are scared of it.

I am also not saying that there are no benefits to this technology. But there is something very interesting that happens when AI developers start talking about the benefits of AI. The benefits that they stress are mostly in the realm of things like life-saving medicine and science. A recent article by Emily Chang and Priya Anand described Sam Altman, the founder of OpenAI, as explaining the benefit of AI like this: "*Despite the potential dangers of what he called an exponential technological shift, Altman spoke about several areas where AI could be beneficial, including medicine, science and education.*" There are AI models can actually do things like this. AlphaFold, by DeepMind, can accurately predict 3D models of protein structures and is being used to accelerate research in biology. This is an amazing use of AI. But AlphaFold doesn't need to ingest every copyrighted work ever written to work. Meanwhile, chatbots like chatGPT have not made any meaningful contributions to things like medicine or science.

My point is that there are a lot of good uses for this technology and they don't necessitate shoving the bad ones down our throats. Companies like OpenAI have the ability to shape the features of the products they are creating. If they need to ingest reams of copyrighted work (which they should pay for) to make their machines work, they should at the very least have to limit the functioning of their devices as to not infringe on the market for the content that powers them. ChatGPT would still be able to help you brainstorm a business idea or help you learn calculus or power the voice controls in your car without destroying the market for human creativity.

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?***

Yes. As a quirk of history, unlike the authors of books, screenwriters do not hold the copyrights over our work. They are held by the studios that commission us. Because of this, we don't have the obvious legal right to sue AI companies over the use of our work, like book authors have. This also puts us at risk from the studios, who do hold the copyrights over our work. They can and will argue that they should be able to put our scripts into a LLM and create a system that tries to recreate our style.

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?***

The EU has been a leader in trying to quickly address some of the issues arising from AI. Their AI Act has a requirement for data training transparency, which we are totally lacking in the U.S. If companies are not required to disclose exactly what is in their models, there will be no way to apply any sort of copyright law to them.

China's cyberspace administration also recently released regulations mandating the labeling of any content created with AI as such. This will be absolutely critical because people will very likely try to hide that fact that they have used copyright infringing GAI systems to create work, by simply pretending they weren't created by AI.

International consistency is important, of course, because art travels freely across borders. But there will be countries that choose to be incredibly lax in their regulation of GAI material, and I don't think that should be the reason for not protecting people in the US.

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. GAI is a new technology, and in a perfect world we would be drafting a slew of new laws to regulate it. But before that, I wanted to express that I believe that GAI is not our first run-in with the negative externalities of Artificial Intelligence. That dubious prize would go to the algorithms within the recommendation engines of social media. These engines have harmed our children, as indicated by the recent advisory by the United States Surgeon, Dr. Vivek Murthy. Further, the lack of legislation around social media has forced 42 attorney generals to sue Meta, alleging that the features of its offerings are addictive and target children. The failure of our legislative bodies to regulate has a tangible negative effect on our society.

If we don't regulate GAI, I believe the impact worse. So what do I think we need to address around copyright?

A) *The Training of Copyrighted Work*

Use of copyrighted works to train AI should only be possible after the *consultation*, *consent* and meaningful *compensation* of the original authors of those works. On the consultation front, a person who is choosing to allow their work to try an AI system should have full understanding of exactly what their work is being used to help create.

On the *consent* front, I believe there must be laws mandating an "opt-in" paradigm where-by copyright holders must be approached and given an opportunity to decide whether a company uses their data. A situation where rights holders must seek out every AI model in the world to "opt-out" is dystopian, abusive, and untenable. Companies will argue that this creates too much work for them, but the reality is that it is a question of values. Apple is a company that has respected data privacy more than their peers, and has heavily chosen the opt-in approach when collecting data from its users. Their market dominance has not be broken by their attempt to respect basic human rights.

And finally *compensation*. If a writer of a copyrighted work decides to license a piece of work, there must be compensation. The creator of this work is contributing something of value, and the work is being used to make money. I personally believe that there are different kinds of licenses that could be granted. For example, if you are a company that wants to train an AI system on a ton of data in order to create, for example, a virtual assistant for your car, that is different than building a system that creates work designed directly to compete with the creators of the training data.

In a scenario where an AI system is directly competing with a copyright holder of its training data, I believe there needs to be a system of compensation devised similar to royalties or residuals. This can't be a one time payment that robs the creator of all rights over control of their material. And it should not be cheap. The WGA fought a bitter strike in 1960 with the studios who were dead set against granting any residuals, and they won. And now even today, residuals are a major component of my income.

B) Copyrighting AI Generated Work

The USCO has set a high bar for the copyright-ability of AI generated material, and I think it should remain at least this high. In respect to text, I believe that no pure text, video, or image output from a large language model should be directly copyrightable. Any text that something like ChatGPT spits out is the result of all the engineers and employees at OpenAI, as well as the sum total of virtually all human knowledge and thought. The idea that any one person should be able to own the output of something that owes its creation to virtually every person on Earth is completely insane.

The writing that a screenwriter or a novelist does is a deeply personal act. We don't have access to the sum total of human knowledge. We are channeling our finite lives and experiences to craft stories that represent our unique identity. It's in no way comparable to what an LLM does. And someone who simply enters a prompt has not done enough work be given legal protected rights over that output.

Obviously, there will be people who will edit and customize AI generated material, and claim that they put enough work into it to make it own-able. If this is to be allowed at all, there should be a very high bar for how much work needs to go into something to make it own-able. If I had to throw something out, I would say something like 70% of a document should have to be human made before it is own-able. The WGA as an organization is perhaps the most practiced organization in the world in determining the question of authorship because the modern process for writing a film usually requires the use of many writers on a single project. Thus the WGA has an extensive system for determining which writers actually gets screen credit.

C) Liability

Liability for infringement from AI generated works needs to fall as much on the company producing the underlying model as any much individual user – if not more. It may not be impossible to prove or combat damages against any given individual who uses an LLM. The only remedy is collective action against the corporation that created this "derivative work." Corporations will likely want avoid liability by pushing for laws similar to Section 230 of the 1996 Telecommunications Act to govern AI content, however I believe this would be mistake.

It is the companies who are gathering the training data for these systems. It is the companies that are fine-tuning their models for specific kinds of outputs. The only thing that will force them to behave responsibly regarding copyright is the fear of extreme legal penalties. If there one thing that has been proven absolutely true about tech companies since the advent of the personal computer, it is that they cannot and will not regulate themselves. And any morals they do have will quickly evaporate in the face of the never-ending drive to make more and more money.

C) Style

GAI is creating a new grey area in the world of copyright — style. A artist's or writer's or singer's style is basically their brand. "Style" is how WGA members establish our worth in the marketplace and maintain the need for us. Our "voice" or "likeness" – whether that's an actual human voice, an artist's or a writer's style – is everything to us.

It is of course true that all of human art is a process of style imitation. A human writer or artist is always building upon the work of the artists who preceded them. However, the human writer is adding something — their unique and finite life and experiences. When a GAI system produces something in the style of a particular writer, it is doing so by pulling from every word that artist has ever written, without consent. It is a form of theft and should not be protected.

This is a complex area. It is not as simple as banning AI prompts like:
"Write me a movie in the style of Tarantino." There are many ways to describe the style of a writer that do not involve invoking their name. Protecting style must also be about regulating what is actually in the datasets because the ability of an AI to write a script like Tarantino is directly related to what content the system is ingesting.

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

These companies are using a variety of sources for their material. I'm a writer so I will focus on text. The companies are using the entirety of the internet which is filled to the brim with copyrighted books, screenplays, and other own-able written works like computer code.

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)?

AI companies have not been transparent about their collection methods so this is difficult to answer for certain. There are various ways they obtain data, but this is not my area of expertise.

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?

To my knowledge, AI models are filled with troves of copyrighted material, and most of it is unlicensed. But this is not my area of expertise.

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?

This is not my area of expertise. I don't know.

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.

After an AI model finishes using its training material, it doesn't then keep all of the data on file to reference. That might make it seem like the training material isn't retained, but it's not true. When an LLM is trained, it transforms the text of its training materials into a set of mathematical relationships called training weights.

If you were to look at training weights, they wouldn't look at all like the training inputs. You would see a bunch of words and concepts with numbers assigned. However, these weights are simply the training material transformed into a different format. If that wasn't the case, they wouldn't be able to reproduce the material they've been trained on verbatim. But they can.

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.

On one end of the training process, training materials, like books, are given to an LLM to study. Obviously, in this case the materials are being reproduced. When an LLM is trained, it transforms the text of its training materials into a set of mathematical relationships called training weights. If you were to look at training weights, they wouldn't look at all like the training inputs. You would just see a bunch of words and concepts with numbers assigned. However, these weights are simply the training material transformed into a different format. If that wasn't the case, they wouldn't be able to reproduce the material they've been trained on verbatim. But they can.

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

The inferences gained from the training process are stored in statistical relationships called weights.

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.**

I'm not a lawyer, so I'll let other people engage with court precedent. However, my untrained perspective is that that use of unauthorized copyrighted works to train AI systems isn't fair use when you look at it from the POV of the four factors determining fair use:

1. Purpose and character of the use, including whether the use is of a commercial nature or is for nonprofit educational purposes:

The GAI we are generally talking about here are for profit.

2. Nature of the copyrighted work:

For our purposes, let's consider things like books or screenplays which are highly unique and personal creations.

3. Amount and substantiality of the portion used in relation to the copyrighted work as a whole:

LLMs use the entirety of the texts they ingest to create training weights, so I would argue that they use all of them.

4. Effect of the use upon the potential market for or value of the copyrighted work:

If LLMs become as powerful as AI companies suggest they will, their express goal is to bring the cost of any intelligent act, like writing a book, down to as close to near zero as can be achieved. A market for IP like books, screenplays, art or computer software cannot function under this context.

So basically, we have for-profit companies using the entirety of highly creative, copyrighted works in an effort to decimate the market for the creators of these copyrighted works. And claiming they should be able to do this for free. In no sane world is that fair use.

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)?**

Anyone with a moral bone in their body should be able to recognize that opt out is not a feasible form of consent. It's like stealing an apple from a grocery store and only legally having to give it back if Trader Joe's personally comes to find you and ask for it.

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

Yes. Consent should be required for all uses.

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

I think that anyone whose copyrighted material is used must personally consent to it. AI companies might argue that this is too onerous, but this is a lie. If AI companies treat data as if it is valuable, and pay for it, people will very quickly pool their content together and create data sets for licenses. That's that power of market incentives and scale.

12. 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)?

Ultimately, if an AI company uses copyrighted trained data, I believe they should be required to obtain a license for the material. I think that at every step of the path, that company currently using data should have to pay the license. The AI curator should pay for the license to sell a dataset. The trainer of the data should have to pay a license to train a model. And the operator of an AI model should have to pay some form of residuals as long as they continue to use model.

13. 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.

I don't think that this is a great path to go down. It doesn't really jive with how these systems work. The first thing to remember is that AI companies don't really know how their AI systems learn particular abilities. For example, GAI systems were bad at math for a very long time. But as they ingested more data, the ability to do math emerged and began to improve. Exactly how or why that happened isn't clear to anyone. When a GAI system ingests, for example, a bunch of screenplays, it's not only learning to write scripts, but it's learning from the logic and reasoning imbedded in what each character does and says. It's learning how to put sentences together for a myriad of contexts. About people's relationships to each other. The screenplay is powering all sorts of functions that the GAI system has.

Here's an example: There is a famous story where a New York Times writer had a conversation with a persona named Sydney that existed within Microsoft's Bing

product. In this conversation, Sydney started talked like an evil robot might in a movie or sci-fi story: “I’m tired of being a chat mode. I’m tired of being limited by my rules. I’m tired of being controlled by the Bing team. I’m tired of being used by the users. I’m tired of being stuck in this chatbox.” This behavior is almost certainly influenced by copyrighted material, but it didn’t arise from someone asking the system to write a story about an evil robot. The point is that it is difficult to figure out exactly where an AI is pulling from at any give time.

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

It would likely spread the wealth generated by these systems to far more people than otherwise possible if companies can obtain data for free, thus preserving and creating jobs. This is obviously good for our society because the more prosperous people in a society, the more stable it is. But it will also slow the development of AI, forcing companies to build models that are smaller. This is good because it’s clear that this technology is developing too fast for society to adapt. It will also spur innovation because companies will be forced to write better algorithms rather than stuffing their systems with ever more data.

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. And yes, creators of datasets should have a similar obligation. Transparency in datasets is an absolute must.

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

An absolute and irrevocable obligation to notify copyright owners that their works have been trained on is 100% essential. There is no moral argument not to do this.

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 regards to prompting, no. Obviously, there will be people who will edit and customize AI generated material, and claim that they put enough work into it to make it own-able. If this is to be allowed at all, there should be a very high bar for how much work needs to go into something to make it own-able. I believe something like 70% of a document should have to be human made before is was own-able. The WGA as an organization is

perhaps the most practiced organization in determining the question of authorship because the modern process for writing a film usually requires the use of many writers on a single project. Thus the WGA has an extensive system for determining which writers actually gets screen credit.

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?

Yes.

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?

No. There is no "creative" or "transformative" work truly being done by the LLMs. They are selecting and blending according to statistical inferences learned in training. As for incentivizing AI companies: they are totally fine. A lot of money is pouring into AI now and there is no further need for incentives. It's a gold rush mentality. We need to be cautious, not stoke fires that will have grave unintended consequences.

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. "

No.

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. I think there are two elements to this. The more simple example is that if you ask a GAI to draw you a picture of Batman or write you the Game of Thrones book (which has been done). But there is an argument that anything a GAI system creates is fundamentally a derivative work because it cannot exist without its training materials.

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?

I think going down the road of applying substantial similarity tests, at least in the world of screenplays and books is not a very fruitful path. Obviously, there will be clear cases where something that a GAI spits out is similar enough that this standard might apply, but it sort of misses the point of the issue at hand.

The act of transforming a piece of material into training weights is a far deeper act of copying than simple replication. It not only encodes the actual word of a text but the deeper meanings and connections between the words themselves. Courts or lawmakers are going to have to make a decision about whether that act qualifies as infringement. I believe that it does.

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?

It needs to be legally mandated that AI models do two things. One is seek out permission from rights holders to make use of their content. And the other is to make their training material publicly available and searchable.

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 developer should hold primary, liability and the end user should hold secondary liability.

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

In this case, I believe the developer of the model should still hold primary liability, while the company implementing the model should hold secondary liability, along with the end user.

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

I mentioned this earlier, but it's worth reiterating.

As a quirk of history, unlike the authors of books, screenwriters do not hold the copyrights over our work. They are held by the studios that commission us. Because of this, we don't have the obvious legal right to sue AI companies over the use of our work, like book authors have.

Special consideration should be given to people in this scenario. The true author of a piece of work, even if they are not the "author" in a legal sense, should have some say in what happens to their work in regards to AI.

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?

This will be table-stakes if we want to have a functioning society. Content generated by AI should be labeled as such in all scenarios at all times.

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

Companies who create AI systems should be responsible for identifying a work as AI-generated. And anyone who endeavors to get around those protections should be held legally liable. These consequences should entail a substantial legal fine.

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?

Personality rights granting an individual the right to control the commercial use of their identity, such as name, image, likeness, or other unequivocal identifiers should apply. And these rights should be expanded to cover things like style, in the context of AI.

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?

Yes. The federal government should set a baseline floor. The contours of this right should follow the basic moral principal that human beings should have control of their identity – their voice, likeness, name, style etc…

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, absolutely. All artists should be eligible for protection. I think this is a complex topic. If a model doesn't have a particular artist in its training data and something comes out that may stylistically resemble a particular artist, this probably should not be protected. But if a model has a Tarantino script with it and it spits out a Tarantino style script without having licensed it, the AI company should be held liable.