Fair Use or Foul Theft? Copyright and AI Training

Does the use of society's resources come with an enforceable obligation to give back to society?

Publication note
Originally published in Cubed. Author's copy on Substack.

Does the use of society’s resources come with an enforceable obligation to give back to society?

The issue of AI systems being trained on copyrighted materials without permission has become a pressing concern. Some argue that this practice qualifies as fair use and is a necessary step toward innovation. Others see it as a violation of creator rights and a growing threat to human employment. Existing copyright law, especially when applied through lawsuits, seems ill-suited to resolving the conflict. Perhaps the real issue isn’t infringement of individual works, but the extraction of society’s collective labor on a massive scale. Copyright law does not address that problem, so perhaps the solution isn’t litigation. Instead, we may need new legislation that acknowledges AI’s fundamental debt to the public in a sensible and constructive way. The goal should be to support the growing number of people displaced from jobs by AI, without obstructing AI’s continued development and use.

Note: An earlier version of this article was originally published on Medium.

Training AI Systems

Artificial Intelligence (AI) is a blanket term that encompasses a wide range of technologies, from predicting stock prices to self-driving cars. AI systems don’t come into existence already capable of producing interesting output. Instead, AI systems must be “trained” before they become useful. There are many ways of training an AI system, but most of them involve massive amounts of data in the form of examples. In some cases, these examples are data that don’t belong to anyone, such as weather histories, shapes of protein molecules, or stock prices. However, in many cases the examples are billions of text, image, and audio works previously produced by human creators.

The problem is that this data from human creators is owned by those creators, but in many cases it has been used without explicit permission to train AI systems. Some people think that this use in training is a violation of the owner’s copyrights, while others think that it is fair use. These issues are currently being debated in several court cases. One of the reasons that the problem is so perplexing is that when an AI system is trained on literally trillions of documents, there is no clear way to attribute a measurable amount of the resulting AI system’s value to a specific work, or even a set of works.

Improper Use

There is a second copyright issue that relates to if the output from a Generative AI system could be infringing someone's copyrights. However, I think that this second question is an easy one: just apply the same standard that is used when a human produces something.

For example, the New York Times published a demonstration showing that when an AI system, called Midjourney, was asked to produce "Joaquin Phoenix Joker movie, 2019, screenshot from a movie" and the AI produced an image that clearly infringes on Warner Bros's copyrights. That's not particularly surprising because if a human followed the instructions to create a "Joaquin Phoenix, Joker movie, 2019, screenshot from a movie" image then they would also end up creating an infringing image. If someone uses an AI tool to create an image of copyrighted character then that person is the problem, not the tool they used. As shown in the above figure, a prompt that does not explicitly reference the actor and movie produces a more typical example that is quite different from Joaquin Phoenix's portrayal. Even then, the prompt was essentially a description of the shot from the movie.

The fact that an AI system can generate a good replica of the copyrighted image when given a specific prompt to do so could imply that there is a copy of the copyrighted image somehow encoded in the AI system's trainable parameters, but I am skeptical of that idea. When the full resolution versions of the images are compared side by side, the contrived example differs in many non-trivial ways. While subtle, many of these details are high-level differences, like the pattern on a shirt, or the type of material and wrinkle pattern of a jacket. If a representation of the image were stored then the types of errors would be image-space errors. Conceptual errors indicate that to the extent the image might be stored in the AI system, it is in a descriptive form. Instead of memorizing the image, the network has learned that the character was wearing a green shirt with a pattern.

The technology expert, Gary Markus, has published many other examples, showing that generative AI systems can reproduce copyrighted images and text. He shows that in many cases, a generic prompt, such as "animated sponge," "robot cop," or "video game plumber" results in images portraying copyrighted characters, respectively Sponge Bob Square Pants, Robocop, and Nintendo's Mario. Again, these examples show differences, such Mario with a tool-belt and rolled up pants. While these differences are minor, they again indicate some sort of descriptive representation rather than a copy.

This subtle distinction is important because an AI system that has no knowledge of what copyrighted characters look like would be severely limited. The same would apply to an AI system that is ignorant of other types of copyrighted materials. It seems an almost philosophical question: Is knowing enough to approximately reproduce that something is the same as having a copy of it? I think there is a difference, and will explore this question in a future article.

Before moving on, a final comment about the output of generative AI systems is that the AI systems can be explicitly instructed to make something original and not copy someone else's copyrighted material. Despite mimicking many human qualities, these machines are not people and while an human artist should know not to copy other's work, if we want a machine to know, then we need to explicitly tell it so. It's the user infringing, not the tool.

Fair Use

Getting back to the issue of training on copyrighted materials, the question to address is: If an AI system is trained on copyrighted materials, without permission, does the training process fall under "fair use" and if not, then what damage, if any, has been done to the copyright owners?

The term "fair use" refers to uses of copyrighted material that are considered acceptable, even without payment or permission from the creator. For example, a book reviewer who quotes parts of a book in their review is making fair use of the quoted material. Another example would be a comic who parodies songs.

To determine if something is fair use, there are four factors that should be considered:

- *The purpose and character of the use, including whether the intended use is commercial vs. for nonprofit educational purposes*
- *The nature of the copyrighted work*
- *The amount and significance of the portion used in relation to the entire work*
- *The effect of the use upon the potential market for or value of the original*

Source: https://www.lib.berkeley.edu/research/scholarly-communication/copyright

The issue of the first factor is whether "the material has been used to help create something new or merely copied verbatim into another work." In the case of training an AI model, I think this first factor is pretty clear. As discussed above, an AI model does not include copies of the training examples. Training an AI model is very clearly a highly transformative use. It creates something new that is not just a collection of the training data.

With respect to commercial versus non-commercial, AI models fall into both categories. OpenAI keeps their GPT model a proprietary secret, but Meta releases their Llama model free for the world to use. However, Meta still benefits commercially from Llama, it isn't a charity project. The final uses of both models include many commercial and non-commercial applications.

The second factor excuses use of materials that are more factual than creative. However, given that AI models have been trained on pretty much *everything* on the internet, this second factor is not particularly helpful in this context. The second factor might excuse some instances of using copyrighted material, but there are a vast number of examples that are clearly creative works.

The third factor relates to how much of the copyrighted material is reproduced. In the case of a trained AI model, I think the answer is that copyrighted material is not actually reproduced in the model. The model might be capable of reproducing some material on request, such as the case above of The Joker, but that is an action, separate from training, initiated by a human user of the tool.

The fourth factor is perhaps the most relevant to this discussion. It asks "whether [the] use deprives the copyright owner of income or undermines a new or potential market for the copyrighted work." The interesting part about this final factor is that the market for the a specific copyrighted material that was used for training is untouched by the training. However, the overall markets for the skills and resources that created the copyrighted work are

affected.

To see this distinction, consider the example of the Joker image discussed previously. If someone wants a poster of Joaquin Phoenix as he appeared in The Joker, then an AI model is not a useful alternative. You can't hang an AI model on the wall and it doesn't look like Joaquin Phoenix or the Joker. Of course, you could use an AI model to make an image to print and hang on the wall, but existing copyright law would apply to the produced image much the same as it would apply to downloading an image and printing it without permission. The existence of the AI model doesn't really have an impact on the market for the original image.

Markets of Talents

However, there are markets that are directly and significantly affected by trained AI systems. Specifically in this Joker example, the market for hiring photographer/videographers, the market for lighting designers, the market for touchup artists, costume designers, and any other role that was needed to produce the image. The trained AI system is now able to generate new images without employing all those various skilled people.

This effect on the market for talent, as opposed to an effect on the market for a specific work, is one of the main reasons that many people are concerned about AI. Artists, writers, programmers, and many others earn a living by providing their skills for hire. When AI systems learns to replicate those skills from examples, then the AI system eventually becomes what is essentially a free, or very cheap, alternative to hiring a skilled person.

The positive perspective is that AI is democratizing content creation. If I want to make a movie about homicidal clowns, then I can do that on my own using an AI tool. I don't need a multi-million dollar budget to hire lots of skilled workers.

The negative perspective is that I don't need to hire lots of skilled workers. In my case, I couldn't have hired them because I don't have the budget to do so in the first place. The problem is that the movie producer who would have hired a whole production crew might also now decide to save money by using AI tools instead.

This issue doesn't just apply to image or art. Millions of programmers have produced code that was then used to train AI systems that write code. Millions of authors have written text that now allows AI systems to write text. Millions of musicians have recorded music that was used to train AI systems that now can create music.

A New Problem and Obligation

AI systems themselves don't include copies of copyrighted material and they don't impact the market for specific works directly, which leads me to believe that AI models themselves are not infringing, at least not under current law as I understand it. On the other hand, If someone uses an AI system to replicate someone else's creation then that is already covered by existing law, and the fact that they used an AI tool to do it is irrelevant.

The thing that is not covered under existing law, at least not clearly, is the effect of trained AI models on the market for the skills that were used to produce the training data. People are rightfully worried that their jobs could be replaced by AI systems that were trained, in part, on their own work.

Note that when I say "in part", it really is a very tiny part. OpenAI's GPT 4 was trained on *trillions of words* of text. Even a prolific writer would be just a tiny drop in a huge pool of text. For example, Stephen King, an author

known for his large volume of work, has said that he writes 2000 words a day. That's a lot, for a person. However, it is minuscule in comparison to AI training. Even if he started writing at 10 years of age and lived to be 100, and wrote 2000 words every single day of his life, that would only be 66 million words. That sounds like a lot, but it is less than a hundredth of one percent of what GPT 4 was trained on.

This enormous scale means that no single individual can claim any meaningful ownership of one of these AI systems based on the inclusion of their materials in the training data. However, perhaps society can, and should, claim some ownership. Certainly the companies that put their resources into developing these amazing tools deserve a return on their investment, but they all used massive amounts of publicly available data that should come with some obligation back to society.

Perhaps existing copyright law should not be stretched and distorted to cover these new AI systems. Instead of grouping creators into class-action lawsuits, maybe "the people" should pool together to pass legislation that sensibly taxes the use of AI systems with the rationale being that the taxes are essentially royalty payments for the trillions of unlicensed little bits that were taken from everyone.

Looking Towards a Plausible Solution

It's clear that AI systems are not going away. They will only continue to improve and become more capable. If we're heading toward a future where a majority of jobs are replaced by automated systems, then we have to ask what happens to the people left without employment.

Maybe it's not individual creators who deserve compensation, but the public as a whole. These models have been built on our collective output, everything from eloquent, thoughtful articles to offhand comments on videos. All of us have contributed to training AI, which means its debt is owed to all of us. That debt can be repaid through legislation that sensibly imposes new taxes on AI use.

As more people are displaced from their jobs, we're going to need a way to support them. This could be it.

Note: Although I have worked as an expert in legal cases involving intellectual property, I am a professor of Computer Science not a lawyer. This article is not legal advice, and I would be unqualified to give legal advice. Further, while I believe that my assertions are generally accurate, any specific case will involve its own particular details that could significantly impact conclusions.

About Me: James F. O'Brien is a Professor of Computer Science at the University of California, Berkeley. His research interests include computer graphics, computer animation, artificial intelligence, simulations of physical systems, human perception, rendering, image synthesis, machine learning, virtual reality, digital privacy, and the forensic analysis of images and video.

If you would like to read my other articles, consider subscribing. You can also find me on LinkedIn, Instagram, and at UC Berkeley.