Submitted by Christina Spiesel

U.S. Copyright Office
[Docket No. 2023-6]
Artificial Intelligence and Copyright

COMMENT

I have read through your lengthy list of very real issues that are demanding solution in the Artificial Intelligence and Copyright problems coming before you. I am a scholar, neither a lawyer nor a technologist, but I spend a lot of time in the company of both. My concern is not that you have failed to be quite complete, but that you are missing a statement of values that should guide your decision making. I suggest you look to the history of copyright and fair use for guidance about values.

Copyright

Originally, copyright was included in the Constitution to attract, protect, and provide important skills through knowledgeable and experienced individuals needed for the development of a new country. It encouraged people to find rewards in doing valuable work. Questions of Fair Use were handled largely as informal decisions in the legal system and not made more formal parts of the law until 1976. That the Fair Use Exceptions became a part of the law was to protect essential functions of communication and use in what has become a large and complex society. Uniformity of principles needed to be established so that things like the ability to quote, engage in parody of other creative works were permitted, activities essential new creativity and to promote scholarship and education. All contributed to making a richer and better culture without endangering the interests of individual creators. If anything, by exposing audiences to the subtle contexts of meaning in our cultural products, there would be a greater audience for original works generally. The Sonny Bono Act, extending copyright to seventy years after the death of the author, significantly tilted power away from individual creators toward corporate owners of copyrightable products. It effectively meant that no culture created in my lifetime would ever be available to me in the public domain to use as I saw fit. Public domain was the great balance enacted when copyright was first devised: creators would have a time-limited head start to profit from their labors, but eventually the works would be freely available to the world.

With the growth of the Internet, individuals had access to an enormous library of works – and they also had to confront the difficulty of assuring authenticity. Now, advanced generative technologies are vacuuming up and using all works, without regard to the rights of individual creators and owners – across sectors. GitHub users have found their computer code appropriated despite the contracts appended to their work; writers and artists find their products – the results of their creative labor – taken with no attribution credit and no respect for the legal regime that backed them when the work was published and granted copyright. Version history, so important in creating robust software is also lost for any kind of cultural works. The same is true for much of the training data for artificial intelligence products used without respecting the ownership rights of those whose work has been absorbed into the Internet. Finally, anything anyone puts on the Internet, whether long comment or flip meme or casual observation is subject to being treated as mere data, abrading the distinction between "a work" and a flow of communication. And all are subject to degradation by disinformation.

How can we regulate for the common good?

Systems for tech use of protected work by others is very complicated, I am sure. Some principles seem may be obvious. First, we all benefit if the markets for original works remains robust, which means that they need to be sustained and not competed with. Whatever compensation is devised, that system should lie outside of the individual companies as it would be too easy, too tempting, to cheat on the social contract. We should all remember that bigger is not necessarily better. Having to pay more attention may yield better tools in the long run. Human creators will be more willing to participate if their contributions are recognized and compensated. Lacking that, they should have the means to prevent their work from being included.

More broadly, a large way that we might start to exercise some control is to demand that these tools are not released until there is complete documentation of their creation, including information about the datasets that inform their training. All of this should be in plain English and publicly available. This proposal doesn't make public the software engineering itself; it is making public the working assumptions in building and "educating" the AI products. These should be clearly identified as created by its progenitors, whether individuals or corporations, but not by the named piece of software itself. Copyright should exist for human creators, period. Machine Learning and Artificial Intelligence systems are tools not persons, not even legal fictions of persons (and should not be given that status.) And yes, outputs should be labeled in ways that humans can clearly identify their source.

Fair Use for Big Data

If the goal of government action is to achieve a balance between human creativity and technology companies, which I argue is desirable, we should all remember that bigger is not necessarily better. Having to pay more attention to what is used for a dataset may yield better tools in the long run. Human creators will be more willing to participate if their contributions are recognized and compensated.

Copyright and fair use exceptions must be viewed as balancing mechanisms. There is no balance between large businesses and individual creators. Individuals do not have the resources of hardware, software, size of data sets and sufficient electrical energy to deploy these tools on their own. Human culture has thrived both to serve the interests of the powerful and to resist them. We all benefit from these tensions.

None of these thoughts can be attributed to my affiliations. They are mine alone.

Respectfully submitted,

Christina Spiesel

Sr. Research Scholar in Law, Yale Law School

Affiliated Faculty, Information Society Project, Yale Law School

October 29, 2023