Comment on US Copyrights Office U.S. Copyright Office Notice of Inquiry and Request for Comments on Artificial Intelligence and Copyright Addressing some of the other comments in the inquiry

Commenter: Shuwen Li, Artist and creator
Date: Dec 5, 2023

I would like to take this comment as an opportunity to reply/address my thoughts on a few points made in comments from this same inquiry. Namely the comments made by: Microsoft[1],StabilityAI [2], ,Open AI[3], and Adobe[4] 's comments.

This is the second submission from me, and I would like to focus this comment on addressing some of the points from those comments mentioned above. (And cited with links in the footnotes)

Comment on Microsoft's comment

"Fair Use Permits Use of Copyrighted Materials to Train AI Models "(pg.7 of Microsoft's comment)

- No, I do not believe this is the case. Copyright is established to prevent unfair self-competition. You cannot take someone else's copyrighted work without permission to create a product that competes with said copyright owner directly. And generative AI does exactly that.

"License and Consent Requirements would Limit Competition and Inhibit Technological Progress " (pg.9)

- That is the point of a copyright system, to prevent the exploitation of creators by allowing fair competition (i.e creator's skills) and punishing unfair competition (i.e using other people's work to pass off as your own)

And as for the "Technological Progress being limited", if that technological progress requires exploitative, unethical, illegal, and destructive activities to be made to the workers/ creators and by extension the public, then such technologies shouldn't be considered progress.

Just because something runs in a computer, doesn't make it technological progress. I believe current copyright laws must be enacted and enforced on generative AI, and artists & creator's rights extended to offer more protection against such exploitive technologies.

---

[1] Comment from Microsoft Corporation - https://www.regulations.gov/comment/COLC-2023-0006-8750

[2] Comment from StabilityAI - https://www.regulations.gov/comment/COLC-2023-0006-8664

[3] Comment from OpenAI - https://www.regulations.gov/comment/COLC-2023-0006-8906

[4] Comment from Adobe - https://www.regulations.gov/comment/COLC-2023-0006-8594

"Copyright Protection Should Extend to Creators Using AI" (pg.10)

- **No. A generative AI model is merely a machine software that mixes its training data via prompts and other inputs. While there is a world where those prompts and inputs made by humans themselves could be considered copyrightable, nothing generated from the machines as a result should be copyrightable.**

If a user could use generative AI to make copyrighted outputs, then generative AI would be used as a tool to ignore copyright. In that state, copyright would lose all meaning because anyone could own anything anyone else made just by passing it through a generative AI image-to-image tool. And never mind the already copyrighted data in its training dataset that the models are able to reproduce that the models retain.

There are already over 50 billion images in the LAION dataset (the data set that many image generative-AI models used to train on) Many of which are copyrighted. A poison tree may only bear poison fruit, these models are riddled with copyright issues, allowing these models to copyright their outputs would be like putting these AI companies before all those artists and creators who got their works stolen and who have their works in the training data.

Comment on Stability's comment

"Models learn behaviors, they do not store works. " (pg.13 from stability AI's comment)

- **This is false, generative AI models have repeatedly shown the ability to retain its training data. It's what is known as a lossy compression algorithm that is tuned to mix its training data instead of spitting them out raw. See my** previous comment **for both citations for this claim and images demonstrating this claim.**[5]

Additionally, before the training process can begin, it still takes the gathering of all of the training data before any training may begin. Which is a process that would often take illegal downloads or "scraping" the internet.

Data has been "scraped", data has been stored, and in turn distributed through the models since the models have the capabilities to re-generate that data.

And because those data are often copyrighted, this act becomes a gross violation of copyright laws.

---

[5] Shuwen Li's Comment - https://www.regulations.gov/comment/COLC-2023-0006-7508

Comment on OpenAI's comment

" "memorization" of training data is generally considered by AI developers to be a bug to be corrected, rather than a feature to be pursued."(pg.7 of OpenAI's Comment)

- **This is just a way of saying "We're not sorry for using copyrighted work, we're sorry we got caught."**
Memorization is a "bug" only because it accidentally proves that they're violating copyright. And to avoid getting caught they'll do anything they can to obfuscate the output so it doesn't "look like" they're violating copyright.

"Because our models do not have access to training information after they have learned from it, they are unlikely to duplicate training data in their outputs." (pg.7)

- **It doesn't need access to its training data anymore because the model has stored it in its latent space. Of course you don't need the original training data if you've taken it into your lossy compression algorithm! (See my previous comment for examples of AI storing data[6])**

"Training AI Models Does Not Violate Copyright Law" (pg.11)

- **If the process requires you to illegally acquire data through illegal downloads, "Scraping" the internet, and replaces/devalues the same workers/creators that made those copyrighted contents by competing with them. Then you are violating copyright.**

Comment on Adobe's comment

"Access to data is key" / "AI is only as good as the data it's trained on" (pg.2 of Adobe's Comment)

- **This statement is blatantly admitting Generative AI's reliance on its training data. This is software used to force unfair competition with the copyright owners of its training data. It strictly makes what it stole.**

"New Legal Rights in Age of AI" (pg.8)

- **Adobe talked about how there should be "Rights of liability" and about the"Intent to impersonate".**

But in this section, Adobe only mentioned the liabilities of generative AI's users, not the companies that created them. Their stance is just "We have no responsibilities as creators of this software, it's the user's fault! " When it's them that scraped and took for their training data without permission in the first place!

And the issue with Generative AI was never only about the impersonation aspect, it was also about unfair and illegal competition via copyright violations!

"AI is a transformative technology that has the power to unleash human creativity in new ways, raise
human ingenuity to new levels, and lift our society to unimagined heights. " (pg.8)

---

[6] 6 Shuwen Li's Comment - https://www.regulations.gov/comment/COLC-2023-0006-7508

- **Adobe has said it Themselves, "AI is only as good as the data it's trained on." It by definition could only take you to as high as your training data. To go beyond that would require the incentives of creating new and better artworks/photography/music etc. From actual creators that don't use AI.**

And if copyright for AI is allowed, not only will we not get to unimaginable heights, but rather the creative advancements in fields affected by generative AI would be grind to a halt due to the lack of incentives to create.

Conclusions

With these counterpoints, I hope I have shown how deceptive the entire AI business and by extension the "AI Revolution" Has been. I hope the copyright office is not deceived by the narratives that these companies push. And gives us (the creators) the protections that we need now more than ever.