US Copyright Office - AI and copyright

Submitter: Birgitte Johnsen
Submitting as an individual who works as an artist and illustrator. I also have worked in IT, and therefore have an understanding of some of the technical aspects of the matter.

1. One of biggest issue with AI used to generate imagery and text is how it is replacing the people who actually spend a lot of time and effort to come up with ideas, concepts and explore those in their chosen artistic medium.
We are already seeing how AI has started taking jobs of illustrators and animators. For instance, the intro to Marvel's Secret Invasion was created using AI. There's also the Bradford Literature Festival in the UK that used AI to create marketing material, which was later removed on digital surfaces.

2. Many visual artists already struggle with making ends meet using their artistic skillset. AI means that it's so much easier for someone to go in and create artwork that will look like we made it, and even claim we made it. We are already dealign with issue of art theft. Meaning a lot of artists are spending a lot of time each week both looking to find stolen images and make sure they are reported and taken down and so on. With certain tools we are able to more easily find these images and in turn report them. However, when images have been created to look like they were made by us, or someone claiming we made them, we might struggle to get them taken own. There have also been instances where the finished image, based on a specific artist's images and style ended up being far too close to one of their pieces.
There is also the disturbing thought that AI can be used to continue telling a story, or create an offshoot of a story that has been abandoned. The story that's continued might be entirely different to what the original author/artist would want. And while, yes, both fan art and fan fiction already exists, but the people who are creating these things aren't often able to copy the style or tone of voice, and are likely to therefore mark it as such - fan art or fan fiction. However, the use of AI means the work looks like it was made by the original author/artist, and therefore someone might think they could profit from that in some way.

3. Papers or studies relevant to this notice:
AI Art and its Impact on Artists -
https://dl.acm.org/doi/pdf/10.1145/3600211.3604681

5. It might require more clarification to specify generative AI.
Currently copyright law protects "the fruits of intellectual labour" that "are founded in the creative powers of the mind". AI isn't intellectual, much like any tool like a camera, the software we use and so on. While AI can generate imagery and text, it is only following instructions based on its programming, the keywords provided by the user, and then uses a dataset to generate a thing. The act of the user is simply

creating an order for something they want, and it is being served up — they are acting more like an Art Director than an artist. Art Directors have no claim to the copyright of the illustration they commission, unless they themselves make significant changes to it.

If the user of AI generators have made significant changes to the generated work, then it would be reasonable to give them copyright to some of this.

It's also fair to say that the copyright laws that covers transformative works and collage would also be something that could potentially work with AI generated images that have been heavily changed by the user. However, work purely made by an AI generator should not be covered by copyright, as there is no intellect creating the work.

There is only one exception to where AI generated work could potentially be copyrighted: That is in cases where an artist creates a data model based on their own work/imagery, and no-one else's work. As in those cases the artist already owns all the copyright to the images in the dataset, and is acting as an artist creating work with the AI generator as a tool. However, if it was a user who created the dataset using the work of just one artist, that they have no consent or right to use, then it will not be copyrightable.

6. Training

6.1. The datasets used by the likes of Midjourney appear to be created by third-parties who have mostly just scraped the internet of images.

6.2. To my knowledge NO licensing models are being offered or used. The work has been scraped from the internet without

9. Copyright owners consent

9.1. Consent from copyright owners should be required in all parts of this - both for training AI models and commercial use. Some work that is being used is also licensed work, meaning someone else have paid to use the work for specific things, and for a specific period. Some of these agreements include exclusive licensing deals, not to mention that the content might contain trademarked items, different characters that are specific to a product, brand and so on. These type of works aren't meant for training or any kind of commercial use, and not having a required consent to opt-in.

Though in the case of it not being possible to create an opt-in or opt-out system, then any datasets used to train AI models or used for commercial use should only be using public domain work — and the people who are creating the datasets and the AI generators must understand that public domain is not the same as content put on the internet.

9.2. An opt-out approach does not work. This puts all the onus on the artists to keep their work out of datasets, which disadvantages artists who for some reason do not have access to set these things in action. This could be for any reason such as being locked out of accounts and unable to get them back.

There are also cases where the artists have passed away, and the people in charge of their estate aren't able to make these changes as they do not have the accesses required, even if option out would be what the artist would have wanted. Opt-out should be the default state to all of this.

9.3. There are multiple obstacles when it comes to creating an opt-in system. First off, it might not be possible to create a proper opt-in system, then copyrighted work should simply be kept outside of the datasets, both for training and commercial use.

While setting a "metadata flag" might seem like a feasible solution, there are issues with this. If it's an opt-out version, then social media sites and so on might not honour this or set it on the pages. If this metadata flag is set on the images itself, then the issue also lies with social media platforms stripping the metadata in images (as photos can contain coordinates of where a photo is taken).

The issue also comes when there are bad-actors such as people downloading, selling and re-sharing work, or even claiming it to be their own. They might not be setting the correct flags where they share things. These bad-actors could also upload artist's work to potential opt-in sites or so on.

Instead AI generators should only be trained on work in the public domain. Because if it isn't feasible to get the consent from the copyright owners, then copyrighted work shouldn't be used for training at all, unless the training is done by the copyright owners themselves.

It would also be possible for the companies to license the work from artists, by being directly in contact with them or their representatives.

9.4. If an objection is not honoured, then at the least work should be removed from the dataset to stop it from being used in the future by new users, or users who update their dataset. However, depending on how much work is used, then there might need to be some sort of legal or monitized repercussions.

9.5. If work has been made for hire, or copyright have been sold to a company then the company should be allowed to object its use. For instance, it would make sense for the likes of Disney, Marvel, DC and so on to be able to keep their IP out of data sets and AI training, in the same way as an artist should be allowed to keep their work out.

10. Copyright license

10.1. While it would be very time consuming to contacting artists directly for licensing deals for training AI models would be feasible. This way both artists get paid a license for their work, get a say in what work is licensed, as well as get a better understanding of how AI works. It also means the AI generators get ethically sourced work. If the companies contact the right artists, or their representatives, directly it makes it possible to ensure things are legal, and it can be a better way of eliminating bad-actors.

10.2. If there is a fair way of artists to opt-in where they both decide what work they license, and they get compensation for it, and it's done in such a way to eliminate bad-actors, then it could be very desirable.

10.3. A compulsory licensing regime might not work, as it could potentially impact licensing for products, illustrations, and so on. Additionally, this is an issue that will also impact artists based abroad in unknown ways, considering they might be licensing their work out to US based companies.

11. It would make sense that both the curator of the datasets and the developers should be responsible for securing the appropriate licenses.

14. One issue that isn't covered here is how users can also train their own AI generators, and therefore pick and choose any kind of image they find on the internet themselves. Likewise you have users who might decide to keep using the datasets they have already obtained prior to all this. Meaning there is still a risk of artists work being used without their consent. This also includes users who want to keep using artists names as prompts in order to get the style and imagery they want. Or even entirely training AI generators trained on a specific artist/illustrator without their consent.

16. Artists, regardless of their chosen field and medium deserve to know that their work is used to train AI models, much like people would want to be made aware of their medical records being used in AI training and so on. As our work is licensable to others, it is an overstep of our rights that our work is used in ways we do not agree to. This is even worse considering that datasets might be sold or leased to companies for training AI models, which consist of content that has been scraped from multiple internet sources.

19. The more clarification there is the better. Considering there are court cases where users creating AI generated images and attempting to get copyright to these images, as they were turned down, then it is clear that there needs to be more information to make it clear how human authorship is required. It's also important for it to be clear that comping up with prompts isn't the same as creating something with your own hands.

25. The liability for infringed copyrighted work lies with all parts: The developer of a generative AI model, the developer of the system incorporating that model and end users of the system. Considering that the data model creator put together all the data, and doing so without the consent of the copyright holder of the infringed work. Though the fault also lies with the end user, who might have intentionally created something that infringes copyrighted content, though without knowing anything about the laws. And without the AI generators it would be impossible for this work to be created in the first place. Though in my opinion the developer of the generative AI model and the end user are probably more liable than the developers of the AI generators. That said, the creators of the AI generators could,

and should, make the disclaimers clear that the work created may infringe copyright, and that the end user might be liable.

28. In certain countries advertising are required to enclose, and label the work, where airbrushing has been used. In TV shows it's also not unusual that product placement also has to be disclosed. Not to mention that even on social media it's also encouraged that post that are paid advertising by the creator has to be enclosed to their audience. Similarly there should be a label showing that this content was made with AI, especially as it appears many people might not spot the difference.

28.1. In certain countries advertising are required to enclose, and label the work, where airbrushing has been used. In TV shows it's also not unusual that product placement also has to be disclosed. Not to mention that even on social media it's also encouraged that post that are paid advertising by the creator has to be enclosed to their audience. Similarly there should be a label showing that this content was made with AI, especially as it appears many people might not spot the difference. It is particularly important that it is clear that the art is made by AI in places where it is created in the style of an artist, where they had nothing to do with the image being created.

28.2. There are AI generators out there created to remove watermarks from imagery and so on. These can also be removed easily, which will create some issues if AI generated images are posted to social media, or sold in different ways, be it for a commission, or prints.

32. Yes, there should be protections agains AI systems generating outputted content to imitated the artistic style of a creator. Because this can both confuse audiences, as they might think the work is genuinely made by the artist, regardless of their chosen medium. In the case of artist and illustrators it also means that they risk losing income, as anyone can create the style of work they are known for. This means that it could potentially be possible for publishers, publications and other potential customers would generate the work themselves, instead of actually hiring said artist or illustrator. These generated works might also contain messages that goes against what the artist would want to create, and therefore can have an impact on their reputation, especially when it's not clear that the work wasn't created by them, despite looking like their work.