Veronica Hoffman
Comments on generative AI and copyright
2023-10-28

There is an important difference between traditional “AI” models, which produce output that is significantly different in kind from the input on which it is trained (like content analysis systems used to study and report on sentiment and biases, or classification models that identify faces or eyes in an image to improve a camera’s autofocus), and newer generative “AI” models, which produce output that is intended to be substantially similar in kind to the works on which it was trained – like diffusion models trained on artistic images in order to generate “AI” “art”. The latter are inherently designed to imitate the works on which they are trained, making them essentially amalgamated forgery generators.

My comments will focus on the latter; they may skew towards a focus on image generators, but even where they do, I see much the same concerns with text generators.

The fact that these models are trained on hundreds of millions or billions of works of art does not change the essential imitative character of their output – given the images (with their descriptions) on which they are trained, they compute associations between words and the kinds of images those words describe, and among the different parts of images (like how similar or different in color and brightness adjacent pixels are likely to be). They use these associations basically to predict what an image would look like if it had a description like the prompt a user provides.

These associations, and the generators’ ability to make a prediction that humans would recognize 1) as a coherent image, and 2) as an appropriate image given the prompt, is almost entirely dependent on the images & their descriptions that are used in the model’s training set, in terms of those inputs’ number, diversity, and quality. For example, if a model were trained exclusively on Sunday morning comics, it would be incapable of generating a photorealistic image, and vice versa. If the training set had poor or nonexistent descriptions, it might generate coherent images, but they would bear at best little association with the prompt. And if the model were provided its own outputs as new training data, the quality of its output would in all likelihood quickly deteriorate – which, incidentally, is almost certainly the primary reason any generative “AI” companies support tagging “AI”-generated images as such. They don’t want to pollute their training sets.

Some people try to argue that training “AI” models on innumerable copyrighted works without the consent of those works’ creators should be treated as fair use because “it’s learning like people do” or “it’s just like how human artists are inspired by other artists.”

To be clear, the way an “AI” model is trained is nothing like how humans learn. For instance, while there’s a lot we don’t know about how human brains work, one thing we do know is that our brains are awful at math, especially large-scale mathematical computations, which are the underpinning of how “AI” models work (developers have to find effective ways to convert words and images into numbers and sets of numbers *so that the algorithms can do math on them*).

To be doubly clear, “AI” being trained on truly massive amounts of human art is nothing like a person being inspired by others’ art. The input to a generative “AI” (I do not say its “experience” because it is not capable of experiencing) is just its training set – 100% art with provided descriptions of the art, for an image generator model. For humans, the proportion of our experience that is “looking at art” is vanishingly small compared both to the constant influx of sensory experience of the world around us

and to the constant stream of our inner lives, neither of which even exist for an “AI” model. This is true even for people who make a point of looking at lots of art. We are taking in sensory input every waking moment; we have an endless flow of reactions and feelings and thoughts in response to that input and to our own reactions and feelings and thoughts. Every experience, thought, feeling, sight, smell, *everything* builds on what came before, adds to our mental model of the world, recalls this or that thing from the past, reminds us of someone, and so on. Everything fits into the whole of who we are and how we experience the world. Art responds to that; it is a way of working through and communicating (with and of and to) our experiences and our world. When a work of art inspires us, we *see* it – understand what it’s doing, how it’s communicating, what it is responding to, or sometimes it just *resonates* with us, feels right in a way we don’t have words for. “Usually it seems to the recipient of a truly artistic impression that he knew the thing before but had been unable to express it,” as Tolstoy wrote.

Memorizing [computational adjustments representing] a billion images is not at all like artistic inspiration, and if you want to see that claim demonstrated, try requiring that the training sets for image generators include a proportionate amount (99.99%? more?) of imagery that is not art, but the kind of images that would result from taking unedited stills from the video of everything an ordinary person sees day to day, and see what kind of “art” that model outputs.

Additionally, the level of control over the way their work is used that an artist or writer would cede by choosing to allow generative “AI” models to be trained on it is for practical purposes, total. The control that we lose when our work is scraped and included without consent or compensation, and in disregard of our license terms, is total. The companies producing these models can be expected to strongly resist any pressure to take down a model, remove infringing images from the training set, and retrain the model in response to something like a DMCA takedown notice (if there were something like that for these cases), due to the massive time and resource expenses of doing so, and there is unlikely to be any way to “undo” output from use of the model while it was infringing, so inclusion in a model via its training set is likely to be effectively perpetual. Ceding rights on terms like these amounts to the most liberal (and most expensive) sort of license, and so far the “AI” companies have thumbed their noses at the notion of consent or authorization and taken, wholesale, the work of millions of artists and writers with no compensation at all, let alone compensation appropriate to the usage rights to which they are trying to help themselves.

I would urge the following.

To be permitted to include copyrighted work in a training set, it is essential that companies be **required** to have fully voluntary, **opt-in**, consent from the creator of the work for each work in question and for each model they wish to train, and that they provide appropriate compensation on terms that the creator finds agreeable. This consent must not be a requirement to use or contribute to other services the company offers (e.g., a stock image service), especially but not exclusively for existing contributors, since that would render the “consent” involuntary and therefore not consent. They must have reasonable evidence that the person submitting a work for inclusion is in fact its creator and copyright holder (we all know there is an endless supply of people who steal and claim credit for others’ work on the internet, and we must not permit companies to feign ignorance of this to their profit or convenience). There should be well defined procedures to remove unauthorized work in a timely manner from models when notified of it, to compensate the person whose rights were violated, and to prevent further use of iterations of the model which were trained on unauthorized work. Regulations must be designed and enforced in such a way that it costs companies more to include copyrighted work in a training set without the creator’s consent (even for small-time, individual creators – not just big studios like Disney) than to rectify the situation, or they will never respect individuals’ copyright –

companies respond to incentives and will do whatever helps their bottom line the most.

If the above seems too onerous for the companies building these models, then they should simply not use copyrighted work. If they are allowed to continue pretending that their wholesale grab at essentially all art and writing in existence for their own and their customers' commercial gain at the expense of individual human artists and writers the world over is "fair use", then copyright as we know it is dead.