

How AI generated art ties to creative theft

There's a constant debate happening online between creative communities and those championing AI tools like LLM's and Diffusion models. In this I'm hoping to add some clarity to the artist perspective when it comes to AI art generators. I'm also hoping that this is a decent overview of the situation that can possibly be referenced when trying to dispel the misinformation surrounding artists. There's this idea that because artists work in creative fields that they lack intelligence or deductive reasoning, however this is far from the truth. I've seen a large effort from artists trying to learn about this technology, often citing write ups from professionals working directly in the AI field. While we are often not directly tied to the field itself, we are also no strangers to research and technology as it often plays an important role in our careers.

An argument that often comes up is that AI art is theft. AI art hobbyists tend to become defensive, many of them are simply enjoying a tool that makes them feel like they are able to express their own creativity or create things they could only dream of before. There are also many often involved in tuning their own models to serve their own individual project needs. It should be acknowledged that there is a big difference between someone typing a prompt into an out of the box art generator like midjourney and someone taking time to build out a custom stable diffusion based model. However this isn't about individual harm, the argument is not that generative AI rips off or replicates individual artists, though in some cases it directly is & can do it. It is about the fact that artist data is needed to build AI art tools & make them viable while also threatening to devalue & displace the art industry. This includes industries such as - illustration, 3D artists, photography, UI art, web art, animation, clothing design, fonts, stock imagery, cinema, vfx and many more that I'm sure I've forgotten to list. This also includes non artistic images, including photos of family, friends, events and places posted all over social media. In many cases these images have been ingested and trained on indiscriminately. There have been some safety that has gone into curating what is being trained on, but it is impossible to properly curate almost 6 billion images & growing.

To put it simply, just because someone posts something on the internet doesn't mean it's automatically free or legal to for anyone to grab & use to build out & train a product. There are many different types of art careers, their efforts should not be considered the fuel that makes the machine run. Especially in a capitalist society. What you are witnessing online with the reaction from artists is self preservation & a push back to a perceived existential threat for their careers and craft.

What typically gets offered as a solution by these AI companies is the ability to opt out. Tick a box and we won't use your Data. This is something the creative community has often pushed back on with the simple question. Why does the responsibility fall on us? Should you not need to reach out and ask permission to use our data. Artists are in favor of consent based opt in. If a company asks and someone agrees and is fine with it, then there's no issue. The big problem with opt out is that it doesn't work. In many cases it is up to the user to opt out every single image they have made on every platform it has been shared on. This also ignores the larger issue of someone simply uploading their work under different parameters. While an artist could perfectly opt every piece they make out of these systems, there's nothing stopping someone else from uploading those same pieces and having them be trained on. The second problem is the black box nature of these AI models. Once they have learned what they need, they no longer have a reliance on direct data and it is near impossible to check what has been

used in training aside from a proper audit that checks through a companies un-altered training databases. Models continue to improve and be iterated on. A model 10 years from now might only need 1000 images to be functional but this is all thanks to the foundational training from the original abundance of data.

The argument of human learning often comes up. There's a large difference between the human brain & machine learning. The biggest difference being all the human variables, reasoning & limitations. A human will never be capable of 100% reproducing or memorizing something down to the minute brushstroke. A machine on the other hand is capable of perfect pattern memorization. Should its aim be to create a direct copy of something it is capable of this. Raw data is being compressed & specific data relied on depending on the model or output. An AI will never create anime without ever having been exposed to anime data. A human on the other hand might reasonably stumble upon a style without ever having seen it. Humans have lived experiences, emotions, trauma, interests, perception, preferences and most of all reasoning. There's an individuality in the artistic process that makes it so that no two artists will ever create art 100% the same way. Their influences may be similar but there is always an abundance of unique human variables behind why their art will look the way that it does.

Style is not the problem. Many often get confused and think artists are trying to protect them. A style cannot be copyrighted and is open to anyone to replicate, be inspired by and created derivatives of. What is not open however is the process of taking images to create a product that replicates these styles. One of the most famous examples in this debate is Greg Rutkowski. His artwork has been trained on and used as a prompt 100's of thousands of times to the point where it has affected his work & his name/brand/business. While there would be no issue with a human being influenced by his work, spending years of their life learning how to replicate it and producing Greg Rutkowski derivatives. What would not be okay is directly copying his work or putting his name on your work. A more complicated issue is when you get into commercialized products. Humans building a product or production could reasonably have some of his work as inspiration in what they are making to a point. In most cases the company would be required to license his name & images to legally use his work in their product. This has been an agreed upon framework in creative licensing for a very long time. AI companies have completely decided that they don't need to follow those same rules because the data is compressed into a mathematical equation. One of the legal arguments was that these tools had rights to do this because they were open source & nonprofit. We all know this to be false especially when looking at the valuations & investor support. Billionaires & Millionaires are not jumping into this technology out of the goodness of their heart.

Beyond this there is not much issue with the technology itself, how it works and what it does. Someone who creates or licenses all of their data & codes & trains their own AI model would be seen as fair with no issues. The issues are the foundational models & the AI companies that feel entitled to take and use whatever they want. While it may seem insignificant that an artists work may only be a drop in the bucket with regards to all the data that's been used in training, because of how weighting works, we know that these machines would not be able to reasonably create something similar to their work without having trained on this influence of data. Fine tuning & training your own model off something that has already been pre trained with the internet's data is where the notion of theft comes in. Your model is seen as being built from something that is rotted at the root.