

Devin Aldrich

1. The core benefit that I see in AI is increased accessibility to the creative sphere. many people live their lives without interfacing with art on account of the high bar of time put into many skills in order to enter, while others are stopped before they even pass the starting line because they simply feel they are not creative. AI as a tool removes both of these barriers by allowing direct conversion of an idea to a physical representation that they can then iterate upon, helping reveal to them their innate creativity as an authorial voice and the fundamental joy of making a piece of art come to life. AI is also phenomenal for generating reference material that has never, and possibly may never, be able to exist otherwise, allowing those with prior artistic skills to find new and unique ideas that can further the sphere in ways that would otherwise take years to reach.

however, I worry about AI for multiple reasons. One of these is that people do not learn how to create art through their own agency by using it; should the AI service be taken down at any point, all of their ability to make art disappears, a situation that could cripple many peoples' creative expression and cause them to stop altogether. The current culture of AI also seems to be full of disdain for artistry to the point of actively desiring to hurt them financially, which I feel is a worrying trend. A more passive example of this is this post (<https://twitter.com/yeeniebeans/status/1637095266770665472>), which highlights a reddit post encouraging people to buy sketches and then color them in with AI, saving them money. This seems like a reasonable use of the technology without context... however, artists often sell their colored pieces at a higher price in order to sell their sketches at a loss, because public perception feels an 'unfinished piece' should cost less, unaware of the immense effort that sketching entails compared to shading for most artists... and, naturally, this affects independent artists overwhelmingly more than large companies. This gradient of cluelessness to active maliciousness is expanded on extensively in this video (<https://www.youtube.com/watch?v=9xJCzKdPyCo>) and paints a picture of a subculture that I'm hesitant to ever support. My heart wishes for AI to be a force which binds us together with empathy and understanding, but it seems from the outset that its first order of business is to attempt to drive us apart with indifference and ignorance.

2. I am an independent game developer; my jobs as of writing currently range from Programming, to Game Design, to light 3D Modelling, Rigging, and Texturing duties, as well as User Interface and User Experience. The primary issue that AI has for my field is that it appears to solve problems without actually solving them. If you imagine NASA constructing a Mars Rover, the last thing they would want to add to it is a mysterious box that behaves as desired somewhat consistently, but will also commonly do something unprecedented and dangerous in an uncontrollable fashion. This is what AI is to a game; code which was not designed, is immensely costly or sometimes even impossible to fix, and has a high chance of jeopardizing the user's experience, with little in the way of controlling it. Companies will inevitably incorporate more and more AI into their titles because they optimize for continued profit growth, but I myself will have a mistrust for AI usage in video games until Explainability in AI becomes a more fully matured and realized field of study.

3., 4. I apologize; I do not know of any papers that address these issues, nor of other countries' legislative decisions pertaining to AI.

5. I feel that new legislation is warranted, and that it should classify AI-generated works (those with little or no human augmentation beyond prompting) as uncopyrightable, while AI-assisted works (those with large human augmentation) are differentiated as copyrightable. Regardless of case, the AI functions as an artist, while, depending on the level of augmentation, the human acts as either a commissioner, exclusively relaying ideas to the AI, or a collaborating artist, contributing notable effort and care to the work.

6., 6.1. Oftentimes the works are taken directly from the internet, in a web scraper-like fashion. As a result, copyrighted works on the internet tend to end up in datasets; anything from works of acting journalists to the galleries of active artists. Commonly, they don't filter out these copyrighted works, primarily because they can't differentiate them algorithmically, and also because the design of the AI incentivizes them to have as much data as possible, in order to yield more satisfactory results. The most a dataset scraper attempts to filter out is images produced by

itself and other AI, which are tagged with an invisible watermark for this purpose.

6.2., 6.3. I have heard little to no news of AI companies requesting licenses to use copyrighted works. As for public works, due to the way the companies gather data for AI, public domain works inevitably end up in datasets.

6.4. I'm unfamiliar with AI companies' particular practices, however training data is commonly kept in order to retrain the AI. The process is wrought with issues and commonly requires tweaking and retraining in order to attempt to iterate on the AI's capability. I doubt that the datasets are removed after an AI's creation.

7., 7.1., 7.2. AI is fundamentally a program which takes a scatterplot of data and produces, effectively, a line-of-best-fit. The process is significantly more complicated than just that, but that is the basic principle underlying the process. What the AI creates is generally considered 'of its own design', as it does not store the data within itself, simply a function that it made to plot it. Oftentimes, this function is stored as weights and biases in the neurons of the model; effectively tuning parameters determining when it fires. When you then ask the AI to produce something, it takes in your input, treats it like a point in space, and finds the place on its line-of-best-fit function which is closest to it.

7.3. Due to the AI not containing the copyrighted works inside of itself, it is impossible with current technology to make it 'unlearn' copyrighted works; The AI would need to be retrained entirely from the beginning. For smaller AIs, this could be economically feasible, however for larger AIs like the LLMs produced by OpenAI, I could see the potential expenses becoming prohibitive.

7.4. Due to the AI not containing the copyrighted works inside of itself, it is, generally, impossible to preemptively determine if it is trained on copyrighted works. There are special cases such as overfit AIs which may produce specific copyrighted works consistently, but those are often seen as glitches to be dealt with due to their penchant for not being adaptive.

8. AI in law is definitely going to be complicated. However, I feel that AI-generated works would likely have to be considered derived works, as they are completely, wholly reliant on the works they are given, and on no ideas of their own. As a result, transformativeness and intent of creation would be large factors in determining whether an AI work is infringing copyright.

8.1. I feel that purpose and character should be evaluated at two points: first, the output (what the purpose and character of the act of using the AI is), and secondly, the input (what the purpose and character of using the copyrighted AI training data is). An AI can be made specifically with the intent to infringe copyright, but an otherwise fair-use abiding AI can also be misused by a user to create copyright-infringing material. Both of these, I feel, are important to determining who is at fault, and whether it is infringement.

8.2. If there are already laws which handle cases of being a distributor of copyrighted data, I would suggest those laws be applied. Otherwise, as a distributor of data for AI datasets, they remain an essential part of the process of creating a functioning AI, and I would suggest evaluating their intent as a service to determine whether they themselves are infringing copyright.

8.3. Noncommercial datasets which are later adapted for commercial use should necessarily be re-evaluated for infringement, as they may have exited the conditions of fair-use along the way. Noncommercial datasets being used by commercial AI models, however, should not need reevaluation, as their own goals have not changed. If the AI model infringes copyright by using a noncommercial dataset, the AI model is the one which is at fault, not the dataset.

8.4. The quantity of data used to train AI models can vary wildly; some datasets may be small (and will likely produce relatively poor AI), while others may be incomprehensibly large. I do not feel that the volume of material should not affect the fair-use analysis.

8.5. In the case of independent artists, the effect of AI used to imitate their work can potentially be career-ending. An AI can produce works that look like theirs, at acceptable quality, at a rate which vastly outpaces them. In addition, while AI can emulate an artist's work, it struggles immensely to concisely produce an exact piece that the artist has actually made. As a result, I feel it should be evaluated against their body of work, with closeness to specified works constituting as particularly compelling evidence where applicable.

9., 9.1. I feel that commercial datasets should have to have copyright owners' affirmative consent before adding their works, whereas noncommercial datasets should operate based on an opt-out system. My hypothesis is that this type of setup would help untangle legal gordian knots without handicapping the usage and progress of AI made for research and fair-use purposes.

9.2. I don't know of any tools that would help with this process off the top of my head. However, I feel the ideal process would be to tag works in datasets with their artists' various names and pseudonyms, and allow them to be publicly-accessible for the purpose of being searched by anyone who would like to be part of the opt-out process. After finding their works, they can request specific works to be taken down, albeit with some verification that they are, indeed, the artist or one of the piece's contributing artists. There should also be a way for artists to add their work back into the dataset after it is removed, as a mechanism to handle false-positives in verification, as well as allow for them to change their mind if they so choose. Systems where an agreement of compensation for continued use may also be a useful hash-out, however I feel that companies would be reluctant to do so compared to opt-in and opt-out measures.

9.3. For datasets established for noncommercial purposes of research and the progress of AI as a field, I feel that it may be unnecessarily stunting to their efforts to require affirmative consent, in addition to their use case likely being well within fair-use bounds. For datasets established for commercial purpose, however, I feel that it should be a requirement to have at least some form of consent-managing measure, as the dataset is actively making money off of their work. While I can't speak on the legal obstacles, the technical obstacles are likely akin to database management, which many companies are already intimately familiar with, and as such I don't worry about whether it is possible from a technical perspective. In terms of the hurdles of affirmative consent, they are still able to use public domain materials, of which there are a wide variety, even if said variety isn't as wide as copyrighted works. Some may argue that opt-in is infeasible because you would have to contact hundreds to thousands of artists, but this can be circumvented with the aforementioned database management field's paradigms, in the form of simply letting the artists who would like to be a part of the dataset come to them.

9.4. As far as I can tell, currently existing remedies for infringement are enough to handle AI-centric infringements.

9.5. I feel that if the human creator does not own the copyright, they should not have the right to object to the AI model being trained on them. They already have their own usable axis of agency, in the form of informing the relevant copyright holders of the potential infringement.

10. When it comes to companies' copyrights, I feel their common licensing rules should continue to apply. However, for independent artists, I feel that the aforementioned opt-in setup, wherein artists come to the dataset in order to consent, may be a useful stand-in as an 'implicit license' for artists who haven't set up their own licensing terms, in order to smooth the process. Revocation of this 'implicit license' should also be as easy and painless as granting it, via an implementation of the aforementioned opt-out system.

10.1. If datasets contact a specific subset of artists that they would like to have in them, rather than every possible artist, it can be very feasible for them to hash out direct licensing agreements. If they attempt to contact every possible artist, though, I could see that becoming infeasible quickly.

10.2., 10.3., 10.4. I don't know of any voluntary collective licensing schemes, extended collective licensing schemes, or compulsory licensing schemes. However, I feel the collective licensing schemes may be more

appropriate for the situation; as a compulsory licensing scheme would likely cause the issue to leak back in due to lack of understanding of the fact that there even *is* a license, and where to go to be able to opt out of it. Outside of that, though, I feel that the voluntary collective licensing scheme makes a good opt-in system, whereas the extended collective licensing scheme makes a good opt-out system.

10.5. Yes, I feel licensing regimes should vary based on the type of work at issue in order to better serve the individual types of artists more specifically, as well as be more recognizable as particular regimes to handle their licensing with.

11. I feel that, in the case of who is responsible of securing them, databases are the ones who should secure the license first, and otherwise in special cases the AI trainer or the AI user should secure the license (for example, if they wish to emulate an artist's work with an AI whose dataset doesn't already contain it). As for the obstacles, I can't speak on the legal side, but, should any of the systems previously discussed be implemented, I don't see much struggle from the practical side of things.

12. It is neither possible nor feasible with current technology to determine the degree to which a particular work contributes to a particular output of the AI due to the aforementioned line-of-best-fit explanation. The way that each piece contributes to an AI's learning is nontrivial to discern, as well as time-sensitive, as putting the piece as the first one the AI consumes causes it to have a different contribution compared to if it were placed later along the training cycle, and not in a predictable way.

13. I don't know. However, I do hope that the transparency and boundaries established by a licensing system or other way of managing artist consent would help AI fill a healthier economic niche than the one it's currently in by establishing a bridge of trust.

14. As far as I can tell, I've already gone over all of the factors that I believe to be relevant.

15., 15.1., 15.2., 15.3. Any training dataset used for AI, be it a dataset loaning out to various AI or an in-house solution created by an AI company, should have detailed recordkeeping for what is in it. Developers of AI models which are not themselves maintaining a dataset, however, should not need to, and neither should developers using third-party AI systems which are in the same situation. For the purposes of copyright, the only thing that is relevant to the dataset are names and pseudonyms of the artist of a work, for ease of searching; determining things like whether copyright on a piece has expired can be determined via other means. Such records should be accessible to the public, likely in the same database-like context as the opt-out solution I'd discussed prior.

15.4. The impact of the recordkeeping would likely be an increase in amount of memory required to store a dataset, although not necessarily the amount of memory that would be used to train an AI with it. Compared to the stored data, the tagging would likely be negligible in cost compared to the contribution of the actual training data, even in situations where that training data is text. The failure case of this situation is models with input sizes which are small enough to be outpaced by the amount of memory required by their required metadata, however with the rise of larger AI systems, this isn't likely to happen in a way that will cause major issues.

16. If any of the previously discussed systems are implemented, there would be no need for the AI models, nor databases, to proactively notify copyright holders. The implications and action of opt-in, opt-out, and license systems cover all of the cases that those obligations would be used to fill.

17. I don't know of any.

A warning to readers after this point: I am not a lawyer. These implementations should be tested for sanity by one, should they be considered. Still, I'll try my best to answer.

18. If the human has significant contribution to the AI-generated work, i.e. by augmenting it with their own creative skill after the AI has generated it, the human is a collaborator on the project. In the current paradigm, this means that they are the only collaborator capable of carrying the copyright, meaning that it may default to them. If Artificial Intelligences receive rights in the further future, however, the piece would be more appropriately considered a joint work. Selecting what material an AI is trained on or iterating prompts is not enough to consider the AI's work one's own; the AI is still the one doing the creative work, whereas the human is performing a commissioning role.

19. I am unfamiliar with the nuanced details of the copyright act, and, as a result, cannot answer this question.

20. Legal protection specifically for AI-generated works is unnecessary to encourage the development of generative AI technologies and systems, provided that boundaries are set such that it can coexist within the wider creative sphere rather than compete against it. This is what I believe the increase in transparency and agency of artists will help establish.

21. AI-generated works in and of themselves do not promote the progress of science and the useful arts; unfortunately, currently they threaten stagnation of the useful arts, as the consistent overusage of models' default settings in AI works very visibly displays. As a result, I believe that the copyright clause of the US constitution does not protect them.

22. AI-generated works should not be granted exclusive rights. However, AI-assisted works should.

23. The substantial similarity test is likely the most adequate test we have for addressing AI concerns, although intention of the infringer should still be weighed in to determine whether they have accidentally infringed copyright, are in fair-use territory, or have intentionally infringed it.

24. Having the datasets publicly available should help ascertain when something is amiss, however it's understandable that mistakes happen, particularly in the case of individuals managing their own datasets. I don't know how to address this, but as far as I can tell, regular discovery should be able to cover situations like this.

25. The developer of the AI model and the end user should both be evaluated in terms of intent. If the AI model did not intend to infringe, but the end user did, then the end user is the one who is liable, whereas if the end user was unaware of infringement, but the AI model did have the intent, the AI model developer is liable. If both are liable, the end user is liable, whereas the AI model developer is secondarily liable.

25.1. Open-source AI models raise particular considerations, specifically expanding the model from 'developer and end user' to 'developer, provider, and end user'. For example, Say that A creates a model and releases it open-source. B then spins up his own instance of the model, which is then used by C. A is the developer, B is the provider, and C is the end user. If A makes the model with the intent to infringe, at which point B and C unwittingly use it and commit copyright infringement with no intention to, A (the developer) is liable. If A makes the model with no intent to infringe, B then takes the model and tunes it with the intent to infringe, at which point C accidentally commits copyright infringement with no intention to, B (the provider) is liable. If neither A nor B intend infringement, but C commits it, then C is liable. If A and B intended infringement but C didn't, A and B are liable, so on and so forth.

26. I don't know.

27. My main worry is that people will be reluctant to use AI systems at all due to fears of being found liable for being unaware that the AI they used was intended to infringe copyright. That is the most integral nuance to this situation: finding the intent and determining who is at fault, without harming others who are, essentially, innocent bystanders.

28. If AI works continue using the invisible watermark system, governmental

labelling systems will be unnecessary, as there will already be a consistent way to identify them. That being said though, I feel that the law should require some form of identification for AI works, hopefully cementing something like the watermark system, but in practice I feel it should be more flexible and allow any method of identification which is recognizable once focused on and revealed. This should apply to all art generated, specifically at generation, but be subsequently removed if the work becomes assisted. I doubt that any of this is possible to enforce, though.

28.1 Ideally, an online tool could be provided that could allow users to determine quickly whether an image has been generated by AI. The identification is not required constantly; only in specific contexts. I feel, as a result, that the people will be a better judge of when to look than an assigned arbiter.

28.2 As mentioned beforehand, I am skeptical as to the feasibility of a labelling system. Should an AI model developer not set their model up to label their work, dozens of those unlabelled works will be in circulation well before any law can intervene.

29. As mentioned prior, current AI works implement a watermark which they use to avoid other AI-generated works; this can potentially be the basis for a system of identification, by specifically searching for these watermarks (although there is a small deal of nuance to it, as some actual artists may be adopting the watermark in order to keep their works out of training data currently in the absence of legal protection). Regardless of what solution is chosen, a warning: identification of AI art should not be done with AI models trained to identify them. There are many 'AI work-identifying AIs' being sold currently which are woefully incapable of actually identifying AI works, and most often yield false positives that cause unnecessary harm to people otherwise not involved.

30. I do not know of any legal rights which currently apply to AI generated works.

31. I feel that the right of publicity is enough to stave off AI-generated material, and that the right interlock of consent-management systems will handle the situation in a more nuanced, appropriate fashion than a new federal right would.

32. While I do feel that artists need to be protected, I sincerely hope that establishing boundaries as-discussed will be an effective solution. If a specific artistic style-protecting law is needed, I foresee messy precedential consequences on the grounds of what constitutes substantial similarity in style. Artists already fight with each other over things like supposed impersonation socially; I can't imagine the havok that involving the courts will unleash.

33. I don't know, but chances are that it will need to be discussed.

34. Outside of requesting that AI generations aiming to replicate photographic likenesses of real people be more thoroughly scrutinized, I don't feel any issues have been missed.