



Digital Hollywood Focus on AI and Copyright

with

- **Matthew Asbell, Offit Kurman**
- **James Sammataro, Pryor Cashman**
- **Pamela Samuelson, Berkeley Center for Law & Technology; and**
 - **Scott Sholder, Cowan, DeBaets, Abrahams & Sheppard**

**For podcast release
Monday, November 13, 2023**

KENNEALLY: Welcome to the State of Generative AI Law, part of a Digital Hollywood special conference looking at artificial intelligence, ethics, and the law.

I'm Christopher Kenneally, host of the Velocity of Content podcast from Copyright Clearance Center.

The Constitution of the United States outlines the powers and duties of Congress in Article 1. Section 8 enumerates specific responsibilities – to borrow money and to coin money, to establish post offices, and to declare war. Congress is also given authority to promote the progress of science and useful arts by securing for limited times to authors and inventors the exclusive right to their respective writings and discoveries. And ever since 1787, copyright has raced to keep up with innovation and change. While the US Congress still holds the power to legislate on copyright, the last update of the Copyright Act went into force in 1978, 45 years ago, and the last major legislation, the Digital Millennium Copyright Act, was passed in 1998, 25 years ago. In the meantime, the evolution of copyright law has been left not to elected representatives to shape, but to appointed judges, with fascinating, sometimes contradictory results.

So it is with great interest that the AI training data suits now make their way through the courts. Last month, most notably, the Authors Guild, John Grisham, Jodi Picoult, David Baldacci, and George R. R. Martin, along with 13 other authors, filed a class-action suit against OpenAI, the creators of ChatGPT, for copyright infringements of their works of fiction, on behalf of a class of such fiction writers.

Our discussion today will consider not only the copyright questions involved in such a case, but also the implications across all creative industries when AI systems use copyright-protected content in developing so-called generative AI solutions.



I'm joined by a distinguished panel, and I want to welcome them. Matthew Asbell is a principal with Offit Kurman in New York City. Welcome, Matt.

ASBELL: Thanks a lot for having me.

KENNEALLY: James Sammataro is a partner and co-chair of the music group and media and entertainment litigation practice at Pryor Cashman in Miami. Welcome, James.

SAMMATARO: Thank you, Chris. It's a pleasure.

KENNEALLY: Pamela Samuelson is co-director of the Berkeley Center for Law & Technology at Berkeley University (sic). Welcome, Pam.

And Scott Sholder is partner of Cowan, DeBaets, Abrahams & Sheppard in New York City. Welcome, Scott.

SHOLDER: Hi, Chris. Thanks a lot for having me.

KENNEALLY: Well, we're happy you can join us – happy all of you can be here with us. Scott, we want to start with you, because the case that I referred to, the Authors Guild class-action suit which was brought just recently, is one that you are a very big part of. I'd like you to tell us about that complaint and set up for us what it is the Authors Guild is pointing to.

SHOLDER: Just to be clear, it's the Authors Guild among many other authors. This is a class-action on behalf of the defined classes in the complaint, fiction writers within certain parameters. Really, it's not terribly different from some of the other complaints out there, in that one of the major concerns of the authors and the Authors Guild is with respect to the mass copying and use of their works in AI training datasets – in our case, particularly with respect to ChatGPT. Some of the other cases go into other aspects of the technology. We're a bit more focused on this training issue. Obviously, the issues are nuanced, and I will oftentimes throughout this talk refer you to the complaint, because there are certain things that I can't really get into, either strategically or from a legal perspective. Output is relevant in some ways, but our focus here really is on the mass copying, and particularly of fiction works as represented by these authors and the Authors Guild on behalf of a class of similarly situated people.

KENNEALLY: You pointed out, of course, that it is on behalf of fiction authors, and it's an important point to make for this audience that works of fiction are what is so-called rich copyright, thick copyright. Can we talk generally about some of the differences here and



why fiction might make the strongest case of all with regard to these types of uses and the training sets?

SHOLDER: I'm again going to hedge a little bit in terms of getting into the strategy and differentiating different types of works, but I agree with you on the general principle that we're talking about highly creative works. Again, this is not to say that other works are not creative, including nonfiction and other types of writing. But we are talking about dense works that came directly from the minds of the creators. These are worlds that the authors created from their own imaginations, multilayered, deep characters, and complex plots. This is the type of dense, human-generated creativity that is the type of material that an LLM needs in order to be effective and commercially viable.

KENNEALLY: Scott, understanding the limitations you're under in speaking about the particular case, I want to ask you sort of generally about the notion that I raised in my introduction, which is the relationship of copyright and technology and the game of catch-up that copyright always seems to be playing here, and the further complication, which is that Congress has shown no taste for acting and the public has no time to act on issues of copyright, which leaves things to the courts to decide.

SHOLDER: That's right, and that's always been the case, and in my cynical worldview, probably always will be the case, particularly with the Congress that we have now and have had for the last couple of terms.

Yeah, it's a game of catch-up. It's often trying to fit – I won't necessarily say round pegs into square holes, but kind of oval pegs into round holes. It's not quite right in terms of fit. Congress tries to legislate and create federal laws that can potentially be interpreted and stretched to new technologies. But again, we're talking about decades-old laws that could not possibly have foreseen the types of technology we're talking about here. And we've seen the same kind of thing happen with the advent of – really, the proliferation of digital media, social media platforms, in the music industry – well, the entertainment industry in general in terms of file-sharing, the Napster years, music streaming, torrent sites, that kind of thing. It's all an exercise of trying to see how you can fit it into a law that doesn't really quite fit.

I'm slightly maybe a little less cynical or skeptical in terms of what Congress is doing. There's been a lot of talk. I feel like things – for as slow as they move and are continuing to move, they still are moving a little bit faster than they have in the past in terms of people's concerns over social media and content being widely available on the internet and a lot of discussions with the Metas and Googles of the world. There have been hearings. There are task forces. People's minds are on this. And I think part of that might be because there's a concern among a lot of people, not just creative individuals, about to what extent and when are the machines going to replace your job? When people are



personally invested in the outcome of trying to get protective legislation or guardrails out there, they tend to move a little bit faster.

KENNEALLY: Scott Sholder, thank you very much. James Sammataro at Pryor Cashman in Miami, I want to bring you into the conversation here. You are someone who represents and has worked with many in the music industry. What's notable here about the music industry is the lack of lawsuits so far. Yet the industry has taken action with regard to generative AI. Talk about that and some of the differences from text-based works.

SAMMATARO: Yeah, I think where the music industry differs from literary works in particular is it may be able to actually sidestep some of the more complicated copyright issues that are at issue or how are intellectual property laws going to be applied to AI-generated works as a whole. Where the music industry really benefits and differs is it can rely upon state-level publicity laws. It's not an ideal solution.

Just to take one step back for everyone, about half the states in the country have right of publicity statutes. There's a statutory right to protect one's name and likeness. Just from a historical perspective, the right of publicity was really unlocked in 1953, and it was born from the right of privacy. The court effectively then realized that really what we're not talking about so much is the right of privacy, but really instead the right of publicity, which is the other side of the coin. What it is is we want to give individuals, namely celebrities, those with particular attributes, the right to benefit – to commercially exploit and to protect their unique attributes. Again, state law system. Different states protect different things. Not all of them protect voice.

But what you're seeing – kind of the seminal moment here for the music industry of recent times was the fake Drake song that dropped in April of this year. For those who don't know, there was a song that came out – it was released on TikTok by ghostwriter77 (sic), I believe was his name – an unidentified user who actually just recently, as in yesterday, did an article with *Billboard*. But he releases a song called “Heart on My Sleeve” which used AI voice filters to generate a duet between Drake and the artist The Weeknd, both of whom are signed to Universal. The song was an immediate success. It had something like millions of streams within hours. And ultimately, it was taken down.

There are some questions as to whether or not it should have been taken down. One misstep by ghostwriter77 – I guess it was ghostwriter977 – one misstep was that the song actually did embed a sample of a well-known – it was a tag from a well-known producer called Metro Boomin. That actually gave it some credence that it could be taken down through DMCA notice. But lots of questions as to whether or not that impact was a copyright infringement. Really, what there is no question about is that depending on the state, it's protected by a right of publicity. Really at heart, what you're doing is you're



taking a song that used, again, a voice filter of Drake and The Weeknd and imitated their voice with shocking precision, so much so that there was a disbelief that that actually wasn't a real song, hence why it became known as the fake Drake song.

We were talking about how the law is always kind of catching up to technology, which has been the case since the invention of the Gutenberg press. But one of the interesting things is sometimes you have these weird decisions that don't seem all that significant or seminal at the time, and they later become the important bedrock. Within the music business, where most eyes are – there's certainly a keen eye watching these literary cases, particularly the copying mechanism, which is if you're using copyrighted material to train your AI, is that going to be infringement? Is that a fair use? That's a very open issue – very complicated, very nuanced, as Scott said. That's certainly going to impact the music business, and the labels have already indicated that they want some accountability. They want some visibility. There are certain laws that they're suggesting or rules they're suggesting that Congress impose.

KENNEALLY: You mentioned that the music industry has so far laid back as far as lawsuits go, but you're worried that some particular case may wind up with a bad decision for the industry. How would it handle that if something like this would happen.

SAMMATARO: There's always the concern of a bad decision. I would say in the music industry, I think what you've seen is a couple of responses. We had a Hollywood strike where people are worried about losing their jobs. From the artist perspective, they're really worried about losing their voice. Both literally and figuratively, they're worried about losing their voice. The record labels, for their part, have been somewhat aggressive in issuing takedown notices. Certainly, there were takedown notices that were issued with respect to the fake Drake "Heart on My Sleeve" song. There have been other notable songs, by the way, which have caused a similar ruckus – not quite at the same level. But there was a Travis Scott song that turned out not to be a Travis Scott song. There was also a 21 Savage song. So there has been the takedown mechanisms. There's a very watchful eye.

But also, as I mentioned before, there's been a petitioning to Congress. Really, the record labels' perspective, while watching and seeing how things unfold, is they would like three things at their core that they would hope to get outside of a judicial setting, mainly from a legislative setting. One would be a nationwide right of publicity law which would protect voices in all 50 states. As I mentioned, it's very fragmented. There's some gerrymandering. There's some foreign venue choice gaming that one can do depending on the particular statute. So number one – nationwide, federal right of publicity law.



Two, they would like the ability of the copyright owners to see what material has gone into the AI training models. So they would like to know exactly – when you created a fake Drake song, did you go through a Drake catalogue? Did you use every one of Drake’s songs to come up with that sound, to come up with that imitation or voice-alike?

And then third is they want the labeling of AI-generated content. Some of that gets to consumer confusion. Some of it gets down to just protecting the artist’s persona, the most important and identifying component of the artist in many aspects.

KENNEALLY: James Sammataro, thank you very much for that. I want to move on to Matt Asbell in New York City. Matt, welcome back. We’re talking here about various suits. The ones that have gotten the most attention are the suits like the Authors Guild case – I’m calling it the Authors Guild case for brevity – that Scott Sholder is a part of. But there’s another case that’s quite important to all of this, which is the Thaler case, which was a lawsuit brought against the Copyright Office itself. It gets down to what, in fact, does our copyright protect?

KENNEALLY: And the Copyright Office says that anything that is created by an AI technology cannot be copyrighted. I understood you had some thoughts about what the importance of this is and the challenges that may result.

ASBELL: The Copyright Office has been educating the public, reaching out to stakeholders, and there’s been interest for a long time in machine-generated content and what would be protectable. You sort of alluded a little bit earlier to thin versus thick copyrights. I know in the context of photography, this concept of thin copyright had come out because the photographer doesn’t always create everything that’s in the photograph. How is it all laid out? There’s the lighting. There’s only so much that goes into it. But another aspect of what a photographer does, as a good analogy to the generative AI world, is that they use a tool that is not a human tool, and they have some control over that tool, depending on the tool itself, but it’s variable, that control. What they ultimately produce, they own the copyright to. So this tool, this machine, this camera – no one’s saying that the camera owns the photograph. We say that the photographer owns the copyright in the photograph. But they use this tool.

Generative AI is, at least at its essence, looked at in that way. Where we start to get into trouble is we start to say, oh, well, actually, the computer is the author. The Copyright Office does not want to and has not recognized when anything other than a human has created the work. But they do recognize that a tool can be used to create a work, and they’re trying to arrive at what level of control over that tool the human must have for them to be an author for the work to be copyrightable.



And the platforms that utilize the DMCA – the DMCA, the Digital Millennium Copyright Act, exists as a protection for internet service providers, for platforms, so that they can avoid secondary liability, because they have a system on their website to take down infringements relatively quickly. It's a very powerful and useful tool. They often will rely upon these copyright registrations. So if people can obtain a copyright registration, and they can do that even for AI-generated content whether they should or shouldn't have been able to, then there will be assertions under the Digital Millennium Copyright Act on these platforms asking things to be taken down. And the internet service providers, the platform providers that have this content hosted, will be in a more difficult position to judge – what should they do? When do they take it down? Is there some pushback that should be made on the party asserting their copyright to find out whether they really are entitled to that copyright registration and whether they really should have to take down that content?

KENNEALLY: And you pointed out when we chatted about this earlier, Matt, that there's really no mechanism should that assertion be accepted, because up to now, you applied for copyright, you got it. It's been rarely invalidated. There is no process for doing so, really.

ASBELL: The process would really be in the courts, right? If you look in the analogous offices – that is the Patent and Trademark Office, as an example – the Trademark Office has a mechanism to invalidate or even disallow an application from proceeding to registration. Now, more recently, the Patent Office does as well – the patent side of the Patent Office does as well.

The Copyright Office has this sort of isolated registration process. There is pushback. I wouldn't say that you file for a copyright and you just get it. I've experienced, certainly, on behalf of a number of clients things that I thought were perfectly copyrightable that the Copyright Office begged to differ, and I've had those fights. But there's not so much pushback and there's not so much collection of information. And the platform the Copyright Office uses is not so transparent that we as lawyers looking into those cases don't readily have online access to all the details of what was filed and what was said in order to decide that, hey, actually, I think this shouldn't have made it through, and I'm going to challenge it. So I think there's a long way to go from a Copyright Office perspective.

And then how the DMCA will apply in the context of these sort of improperly granted copyrights, assuming they are improperly granted, since it's left up to the internet service providers to evaluate is a matter of, I guess, their risk tolerance. Do they want to be on the receiving end of a secondary liability claim? Maybe they'll just err on the side of accepting that, well, it's registered. I don't want to question that unless the defendant, the party who posted the content, has some ability to assert that this was AI-generated and is not entitled to registration.



KENNEALLY: I think I would be happy to hear what you thought about fair use, because I think the concerns for lawyers like yourself is it's just very unpredictable. Fair use is a case-by-case basis as I understand it. I'm not an attorney, but I hear about this a lot at Copyright Clearance Center. It's up to the judge to decide based on the four factors. What are your thoughts on that?

ASBELL: I think it has been unpredictable in the past. It was already troublesome for people who were planning to use a work to really know and feel confident that they could safely use it. And I think you found – I think maybe Scott or James alluded to Napster and Grokster and those days. In a lot of those types of cases where the law didn't perfectly apply to the technology, the people who developed that technology were relying on their understanding of the law at the time, thinking they were in the clear.

We've had that happen pretty recently in a trademark infringement case regarding non-fungible tokens, the MetaBirkins case, where very similarly, the defendant kind of knew the state of the law. There were all these cases that were previously found to be fair uses, and they should be OK if they operate within these particular confines. But I think even less so today than a year ago or a few years ago, is there any certainty as to what is a fair use under copyright, and what is a fair use under trademark law? People are at a loss, which makes them – I guess it's good they come to us, but our ability to opine on it is really difficult. There's a lot of, well, there's this side and this factor, but then there's this side and this factor. And in the end, oh, it could go either way. I guess the best thing to do is don't ask for permission, and we'll just find out, or the best thing to do is not do it, which I think is a real struggle for those who are creating or using existing content in what they create.

KENNEALLY: Right. Matt Asbell, thank you very much. Pamela Samuelson at the Berkeley Center for Law & Technology, I want to pick up on that, because fair use really is a very important point with you and your own thoughts on these particular cases. As I understand it, based on precedents in law, you think the defendants in many of these AI training cases have a pretty good chance of finding sympathetic judges to their claims of fair use. Tell us about that.

SAMUELSON: Can I make two really quick points about things that Matt said that I think would be helpful to people in the audience? I promise to get back to your question, because it's the one I'm most interested in. But the Copyright Office now requires that when you register something that has AI-generated stuff in it that you basically have to disclose that, and you have to disclaim authorship. That's actually going to be a very tricky thing for them.



A second point that I wanted to make is that the Copyright Office has canceled registrations when they find out that something's AI-generated that they didn't know. Kris Kashtanova made *Zarya of the Dawn*, a kind of comic book story, and didn't reveal that Midjourney was the way that she generated the images, but somebody in the Copyright Office saw on social media that she had used Midjourney and gotten a copyright certificate. So then they went back and they canceled it. If the Copyright Office cancels your registration, that's a pretty big deal.

I think there's going to be a lot of nuance that people are going to have to do. There's been an effort to say, hey, but I used all these different prompts, and therefore I had control over it. And at least the most recent of those cases, I think his name's Jason Allen made this image, won a prize, and he still couldn't get a registration of it, even though he said I used 93 prompts to do it. The Copyright Office said, no, it doesn't work for me. That's actually something – that if you don't disclose it and they find out, you will have committed fraud on the Copyright Office, I think. That's actually going to be something – boy, there's going to be a lot of tussle on that particular issue. Let me just complement what Matt was saying about the copyrightability issues.

KENNEALLY: Pamela Samuelson, tell us about your views on fair use here and why you think many of these cases are going to be found in favor of the defendants.

SAMUELSON: Again, case by case, this presents some novel issues. But certainly, the defendants will be looking and relying pretty heavily on the Google Books cases – *Authors Guild v. Google* and the *Authors Guild v. HathiTrust* case. They're Second Circuit Court of Appeals decisions, and Second Circuit decisions on copyright issues tend to get a lot of deference. In both of those cases, the copying of millions of in-copyright materials for the purpose of essentially engaging in computational uses was held to be fair use. One of the reasons why in the HathiTrust case is that you could use the HathiTrust digital library to search across millions of books to find out which books actually mention this particular historical figure or this particular historical incident. You couldn't get any expression from the books. You could only say here's the book, here's the page on which that particular person is discussed, and then you could go to the physical library and actually check out the book. That was actually helpful.

So I think that because the generative AI software is separate from the training data – it's distinct objects, right? The software basically doesn't embody the expression in these works in a way that we would recognize. Rather, the training essentially disassembles them into component parts, and then those component parts are basically identified through numbers and then essentially used to predict some outcome. But unless the output has got some expression from this particular input, it's hard to say that there's any derivative work created, and also it's hard to say that there's actually any harm to the market for the



original work, because the original work's expression isn't embodied in the model in the way that copyright law has cared about. So I think those are the two cases that will be relied on, and those are cases that will be very relevant to the ingestion claims.

KENNEALLY: Another case that we should bring up briefly is the Betamax case, which as you told me, for the industry, losing that case turned out to be a good thing, and it may be a lesson for us in all of this. Explain that.

SAMUELSON: Universal City Studios and Disney brought a lawsuit against Sony for selling these technologies that allowed people to make copies of television programs, and they claimed grievous losses, enormous losses, as a result of that. And even time-shifting – no, people can't time-shift without me getting some money. So they actually asked for – as a remedy, they wanted Sony to have to essentially recall all five million households with Betamax machines, and they wanted to disembowel the machines so that they could no longer be used to make copies of television programs.

Well, it turns out that because of the installed base of five million at the time of the oral argument – five million households had these machines – it turns out that then there was an installed base of people who wanted to watch movies and watch programs, and the entertainment industry really benefited tremendously by the market for DVDs, for Blu-rays, and so forth. I forget exactly how many billions or trillions that they made from these things, but it turned out pretty good.

KENNEALLY: And the point is that new technologies may have some infringing uses, but they have many others that are non-infringing, and that's an important consideration.

SAMUELSON: I think that also there will probably be some Sony Betamax defenses to some of these claims. Let's say I'm a user and I want to infringe copyright. So I put in some prompts, and I get out something that's like Snoopy or Superman or Mickey Mouse or some other kind of character and that that's an infringement. Well, I may have been the infringer, but did the maker of the generative AI system intend for that to be something that could be done? Probably not. So if a technology has substantial non-infringing uses, and that's certainly going to be what the generative AI companies are going to argue, then they may not be contributorily or vicariously liable, even if the dedicated user who wanted to infringe was able to do that.

KENNEALLY: Pam, if the cases go the way you say they will, which is for the defendants to be able to use their fair use defense successfully in these cases, are jobs at risk?

SAMUELSON: One of the questions that the courts are going to have to confront is whether copyright law is a jobs program. I don't think it is. Goldman Sachs has predicted that 300



million jobs could be displaced by generative AI. The ones that are most at risk are office workers, administrators, lawyers, engineers, and architects, according to Goldman Sachs. So it may be that we're going to have to figure out as a society what we do to keep people meaningfully employed. Now, Goldman Sachs also predicts that generative AI will open up new opportunities, so it will be offset.

And it's certainly the case that in the past, you introduce a disruptive technology, it's going to have some impacts. John Philip Sousa actually was against recorded music, because he thought it would mean that people wouldn't play music anymore, and it would totally change and ruin our society because people wouldn't be playing music together. They'd just listen to this recording, and it was all mechanical, and it was terrible. It turns out that recorded music turns out to be, actually, a pretty cool thing, and lots of creativity in the area.

So again, many, many interesting questions for the courts to opine on. And thanks, Chris, so much for organizing such a good session.

KENNEALLY: Well, I enjoyed it very much, and I want to thank everybody involved – Matt Asbell with Offit Kurman in New York City, James Sammataro with Pryor Cashman in Miami, Pam Samuelson at the Berkeley Center for Law & Technology at Berkeley University (sic), and Scott Sholder, partner with Cowan, DeBaets, Abrahams & Sheppard in New York. Thank you all for joining me today.

Thank you, Victor Harwood, for the invitation to participate in this Digital Hollywood panel. I'm Christopher Kenneally, host of the Velocity of Content podcast from Copyright Clearance Center. That's all for now.

END OF FILE