

**CONGRESSIONAL
INTERNET CAUCUS**
ACADEMY

Intellectual Property and AI:
The Law & Policy Of Human Content



Brandon Butler
Executive Director
Re.Create Coalition



Josh Rogin
Senior Vice President
Motion Picture Association



Nicholas Garcia
Senior Policy Counsel
Public Knowledge

Thursday, October 17th,
2024 at 12 p.m. ET

<https://bit.ly/aiiprayburn>

#IPandAI
@NetCaucusAC

Intellectual Property and AI: The Law & Policy Of Human Content

Thursday, Oct 17, 2024 12:00 pm

SUMMARY KEYWORDS

AI and IP, human content, large language models, copyright, name image likeness, right to publicity, trademark, data collection, training models, output, fair use, robots.txt, licensing deals, public domain, creative works

SPEAKERS

Josh Rogin, Tim Lordan, Cam Madsen, Brandon Butler, Nicholas Garcia

AUDIO

<https://netcaucus.org/audio/2024/20241017aiandip.mp3>

Cam Madsen

Hey guys, Thank you for coming today to this lunch briefing. My name is Cam Madsen. I'm with Chairman Michael McCaul of the Foreign Affairs Committee. He's also the Chair of the Internet Caucus, along with his co chair, Representative Anna Eshoo, over on the Senate side, we have Senator Thune, who's the Chairman over there. Internet Caucus has been holding briefings regularly since about 1996 or so, with a brief little hiccup during the Covid era. But we're back up. We're going today. We have a very good and interesting panel we're going to it's titled The Intellectual Property and AI, the Law and Policy of human content. So this Caucus has been going on since 1996 the Internet was a little thing back then. There were some folks who said, you know, it's, this is a blip that's not going to be that big, really now, and we all understand this. It underpins almost everything in our economy, and the advances in Artificial Intelligence would not be possible without the body of data that is on the Internet

right now. Our moderator today is Tim Horton, and I'm going to turn it over to him, but anyway, thank you for coming today, and hope you take good notes.

Tim Lordan

Thanks, Cam. Thank you, and thank you to Congressman McCaul. Thanks to the Judiciary Committee for letting us use this room, and specifically Congressman Kevin Kiley. So we're super grateful for that. Today's panel, as cam said, we're focusing on Artificial Intelligence and human generated content, because, as Cam said, that human generated content has been building up on the Internet since 1992 probably a little bit before that and that corpus of content that humans have generated and put on the Internet in a variety of different ways, whether it be Reddit posts, whether it be Tiktok videos, or whether it be audio songs or whatever, all that content that's generated by humans has been used to train these large language models, these Artificial Intelligence models, and the term large language models means they're created by training on large language data sets. And so we wanted to do kind of a closer look, maybe a bit of an overview. We we could probably go have one event on like five different issues that we're just going to touch on today, but we're going to try to give an overview of artificial AI and Intellectual Property when it comes to this space. When it comes to Intellectual Property, there's probably, like a several different types of intellectual property that is implicated by Artificial Intelligence. When it comes to creating these models, you first have to the AI companies have to just collect an enormous data set of content and and pull it into the model and train it. These models don't necessarily understand like what a sentence is or what a video is. It basically just predicts the next lot, next thing that's most likely to happen after another thing, like whether it be like a sentence the Internet Caucus, it says the and it says most of the time, Internet comes after a "the", when it comes to these briefings and then Caucus comes after Internet. It's just a predictive assessment by the AI, but being trained on a lot of different uses of that term, so it isn't deterministic, like a calculator. It's more predictive. And then when it comes to Intellectual Property, so you have to collect all that information, you have to train the model, as I just said, like training it against that data set, and then it's the output, the third part is the output. What does it generate? Um whether it be, you know when you go to ChatGPT, and you get an answer in text, or whether its stable diffusion and you get an image, or maybe another one when you get audio that sounds like something. I think the Intellectual Property issue is implicated in all three of those stages, collection, training, and output range from copyright. Now, we're in the Judiciary Committee, copyright we've been, they've been dealing with copyright on the internet forever, there's a long body of copyright issues on the internet. There's also issues related that are relatively new to the internet, which are name, image, and likeness issues of the output, if it sounds like Scarlett Johanssen, is it her name, image, and likeness, or right to publicity? There is no federal right to name, image, and likeness statute. But thank goodness for the Supreme Court and the NCAA, most people today who now watch football kind of have an idea of what name, image, and likeness is. That's how Sheduer Sanders from the University of Colorado Football, can make more money, probably, than in his first year in the NFL, because he can use his name, image, and likeness. So that's, we're thankful to the NCAA for failing to do anything about name, image, and likeness, and the Supreme Court so that everybody can understand what that is. Those are kind of the issues, there's also trademark that has come into play a little but when it comes to Artificial Intelligence in these models, and we might touch on that. But those are kind of the three issues of Intellectual Property, kind of, copyright, name, image, and likeness, right to publicity, trademark, and then the three stages of AI is collection, the training, and the outputs. So, let me just get to the panel. Today we have Brandon Butler, who is down at the end. Brandon Butler

is the Executive Director of the Re:Create Coalition. Josh Rogin is the Senior Vice President for the Motion Picture Association. And then Nick Garcia, to my right, is a Senior Policy Consult at Public Knowledge. So let's kind of, start with the first question, which is, there's a bunch of different pieces of legislation, and we're not gonna go all over them, there's also a ton of issues in the courts. There are lawsuits all over the place, the New York Times had a lawsuit, Sarah Silverman had a lawsuit, there's a dozen or so that are prominent cases. Let me just lay the table, should we be thinking about, I talked about text generation, audio, and then video. Should we think we should be they all the same when it comes to Artificial Intelligence? Or should we think about audio, text, video, audio, video, as separately, or are they all the same when it comes to artificial intelligence policy?

Nick Garcia

I'm happy to start on on that. It's a great question, because everything is getting smashed into one bucket right now. Artificial Intelligence means a lot of different things to a lot of different people, even generative AI and these machine learning technologies is an over broad bucket, in some sense. And so when we're talking about these different forms of media, whether that's text or audio or video or images, I think it is really important to keep in mind that there's different issues here, both in terms of the technology, because there's different models and different training techniques that are used for different things. So the diffusion models used to create image models or multimodal models is different from the transformer technology that's used to create large language models. So there might be different copyright implications from how that technology actually operates, and you also have very different landscapes in terms of the actual content and the rules around that content. Music, for example, has a very different universe of content that exists on the Internet and has very different existing licensing structures than there are for the vast amount of text that exists on the Internet, for example. So all of these are going to raise different copyright issues and different technical issues as well, I think

Tim Lordan

Josh, your member companies are in possession of a few bits of video, being the Motion Pictures studios. How do you look at like the different types of content? And

Josh Rogin

Sure, no, I appreciate it, and appreciate you hosting this event. Is this on? There it goes. So, yeah, I think the better way to look at it is how you started with, kind of the looking at the large language models and looking at what they're, you know, the stage of scraping or inputting data, and then the outputting data. But I do think at the at the beginning of this event, it's, it's, important to lay the table for what's going to be happening in the next year plus, because this is going to be a long road to solving a lot of these questions. So there are a number of cases out there already by authors, by newspaper organizations, by other creators, that have sued these large language models for copyright violations, and some of these cases have already been thrown out. Some of them are on their way to going to trial, and they're going to be a bunch of answers we're getting both on the input side and the output side, so those and eventually that is going to be a it's going to be decisions, you're going to have appeals, you're going to have reversals, and eventually the Supreme Court is going to rule on this. So that could take a year, year and a half, who knows Congress is in the background, watching that and and, you know, it's really hard to legislate on copyright issues as as it is hard to legislate on almost any topic, but copyright is very complicated, and there's a lot of stakeholders. So essentially, Congress is

going to view what how those decisions come out, and whether or not they like them or not, and then that's going to start the congressional process. Do they like where the court landed with copyright? Is copyright fit to purpose for these big AI questions for large language models, or are there other other rights or laws that need to be looked into? And I'll just mention, and I hope we can talk more about it. But you know their proposal, like the COPIED act from Senator Cantwell and Blackburn, which is not copyright law, but it's a bill that would require watermarking or provenance information, and then potentially that provenance information could be used to keep them out of data sets. So that's not copyright law, but that's another way that Congress might respond to these challenges.

Tim Lordan

Great. So, so let's kind of look at the collection side of it. You had mentioned the watermarking to kind of, maybe modulate the collection by different companies. There's also just a huge corpus of content on the Internet. And on Internet headers and on our website, we could put in robots, a robots.txt header to say robots can scrape our site, or no, don't scrape our site and ingest the content. And then there's been cases about whether scraping is allowed and whether that's fair use. And there's cases on whether or not when Google actually or any search engine, whether it be DuckDuckGo or Bing, takes content from the Internet and says, when I type in a search for something it says, "here's what you're looking for," because they've ingested that. They've kind of crawled it, they scraped it, and they give me a snippet of what I'm looking for in my search results. There's a case, a long standing case, an important case on the internet, field, field versus Google. And there's a Fair Use cases. Can maybe, maybe Brandon talk about Fair Use. And from a legal perspective, Fair Use isn't an affirmative right. Law school students would say it's actually a defense. So it isn't like you have this right to fair use. It means that if you get sued, you can, you can say, I have a Fair Use defense. That also means you have to litigate.

Brandon Butler

Oh, you're gonna, you're triggering me now. So yeah, thank you. This is really, this is really important issue, the right of Fair Use, as I would call it, is an important part of copyright. When we talk about copyright, we often talk about, what does copyright protect, right? And the first thing you think of when you think about what copyright protects are the creative works right that copyright attaches to as soon as they are created or fixed in a tangible medium. But copyright protects all kinds of things. Right? Copyright also protects the public domain by ensuring that it is not encumbered by anyone's exclusive control, but it is free and copyright also protects Fair Use. Fair Use is the right that we all have to to use, to copy, to collect and in the case of search engines, right, to crawl the internet and make copies of millions and millions and billions of Internet pages every day in order to create a tool that does something new and different. Right? Right? It doesn't replace those websites. No one goes to Google. Instead, they go to Google to find the website, right? And so when you take information that is part of a copyrighted work and you analyze it, and you create new information, these are the most important websites about a topic, right? That's new information. That's what copyright is all about. Copyright encourages the creation of new information again. One way it does that is by saying, if you create new information and you write it down in a creative way, that's a work that you get to control. But another way that it protects the creation of new information is saying, If you come along later and you discover a way to, again, collect, crawl, process, a bunch of copyright protected works granted right. They are protected by copyright, but what you do at the end of the day is make something new and not a

replacement, then that's a Fair Use, right? And that's solid, solid case law. So we've sort of been down this road, *Field v. Google* was a case, you know, *Perfect 10 v. Amazon*, there was a whole bunch. It was just like, right now, when everyone got wind that this is how Internet search works, along come the lawsuits, right? But that shook out. That shook out multiple, multiple cases. Now, across all you know, both coasts, right? The second circuit, the Ninth Circuit, and in between that say, this kind of massive collection and processing is a Fair Use when what you're doing is creating new information. So that's, that's sort of the status quo

Tim Lordan

In between *Perfect 10* and *Field versus Google* was that use, the use that they claim Fair Use, was that for providing search results and things like that, or is this the different type of use of that content? That's the question. I think,

Brandon Butler

Yeah. So, I mean, there were multiple so the lawsuits sort of covered the water of different things that a search engine does. So there's the collection part. You know, you copied my website without asking. Um. TAnd then there's the presentation of snippets and thumbnail images, right? How do I know if this is the image I'm looking for? Well, the search result would have a little thumbnail, *Perfect 10*. Pornography. Of course, it's always about pornography, *Perfect 10*. A pornographer sued Google and said, "No, you know people, the thumbnails, those belong to us." The court said, "No, that's a that's a way for people to tell is that the image I'm looking for, then they click on that and they go to the place they're trying to get to," right? So we've, we've, we've, we've surveyed a lot of different elements of how this technology works, and it's all been found to be Fair Use. So we've got a pretty strong body of case law for internet search.

Josh Rogin

And I would just say that it's the courts are gonna have to decide whether or not the searching on the Internet is the same as absorbing it all and spitting it back out. So I think we do have case law, but I think this is a novel question. I think this is gonna be, we could have courts going different ways and and it's gonna hard. It's hard to fit this new massive, you know, vacuuming up of data and spitting it out into the current copyright regime, just because it's so more massive and different. So that's what the courts and Congress are gonna have to wrestle

Tim Lordan

So you're saying, you're saying that's a different use than providing search results or thumbnails, taking an entire movie, taking entire, like all of *Reddit* into train a model is a different type of use that's contemplated by *Perfect 10* and *Field*.

Josh Rogin

I think this is a novel, new question before courts. And I I don't know if they're going to follow the line they have followed. It's it's different and and I think judges are being more bold in the tech space in general.

Tim Lordan

There's one thing that I don't quite know, and I don't mean to ask an engineering question, but whether it be, you know, text content, or audio or video, and audio and video, as these things get pulled into these models, there's a question of an engineering question about what is actually being pulled in. Is it the actual, like the, as I said earlier, the AI doesn't think and read and understand. It's looking at just bits of data. And you know, whether it be like, you know, face, like a facial recognition, like, when you go to facial recognition, you don't, does, you don't scan the actual person's face, but it's just data points and like, so it's more, it's not the actual so the engineering course, does anybody know, think have comment on that? Or, Yeah, that might be different.

Nick Garcia

Yeah. I think it's important to to Yeah, have a very clear idea of what we're talking about with these models, because it does affect our copyright analysis. And I think the thing here that's most significant to me is that when these models are operating on all of this copyrighted material that comes in, to your point, Tim, is that what is happening is it's mapping these kind of mathematical relationships between all of the pieces of data. It breaks everything apart, and then it looks at how all of these parts are related to each other. That's the learning process that it's going to is measuring the kind of weights and connections between all of these pieces of data. So if you hear people talking about model weights, what they're talking about are those mathematical relationships that the system is learning between all of the content that it's looking at. And I think this goes to a very fundamental point about copyright, which is what copyright protects, is the creative expression in a particular piece of work, but the underlying facts and data of something that is in a copyrighted work are not protected and never have been, and that's been a core part of a lot of these cases that have existed in the past where people do things like reverse engineer source code in order to create stuff that's compatible with it, you have to copy, in some sense, in order to break something apart and understand it. And that's what these models are doing at a vast, vast scale. And so that's something different from simply making a big copy of everything, putting it in a database and then querying back out little bits and pieces of it. What it's really doing is learning about those relationships, about how do we put together art or text or sentences or stories?

Brandon Butler

And building on that, then you know, the relationships that Nick is describing are facts, right? The when Tim was describing how AI works as a matter of probability, which word is more likely to come next, right? Those are factual inferences from the data. And copyright has never protected facts. You are allowed to derive facts from data sets, and this technology could be really powerful in the research context, right? Universities have been using this kind of technology to derive insights. There was a research team in Canada that sort of saw Covid-19 coming early based on textual analysis of Chinese newspaper headlines, right? You can derive those insights, and those insights are free, and they're extremely valuable. And they are creations of the people who used the tool and developed the tool. And so when we think about, you know, where should the value be? Where is the, Where's, who should control what? That factual information belongs to all of us. And that's an important policy consideration here, yeah

Josh Rogin

And, I just want to respond to that briefly, which is, you know, not everything the large language models are bringing in are facts. So, you know, obviously, if you bring in a New York Times article, there are facts embedded in that, but the how they present it, and, you know, the nuance they put on it isn't a fact. So we just need to be careful about what the data is. And then I just think it's interesting that OpenAI has made deals with publishing companies, Wall Street Journal, others, to allow them to get at this data. Other companies are doing it. They think they have the right under copyright law to do it. OpenAI has taken a different position. So I think there is some disagreements among the players about, you know what's going to happen, what might be allowed under copyright law, what might not be

Tim Lordan

Let's spend a little bit of time on it. I want to go back at some point to robots.txt and what the purpose of copyright law is. But let's open up that, that can of worms about the licensing deals. You probably what we won't know. As everybody just said, we won't know for a couple of years where these cases are going. We could have waited like three years for it all to shake out and do a briefing, but we're a little bit ahead of the game just to get laid down with baseline. But what you are going to be seeing are licensing deals. You've already saw a bunch. Josh just mentioned that deal. I think there was a big deal with News Corporation and their properties, Wall Street Journal, etc. Now there's a question about, does that, and then Google search is the only, only, the only search engine that has a deal with Reddit. I don't know how many of you guys are Redditors or go to Reddit, but like pretty much, if you want to know how to do anything you gotta go to Reddit, and it's so awesome that like it's pulled together for you and just served up. But if they're doing all these licensing deals, if Google's the only one that does a licensing deal with Reddit, doesn't that mean that Reddit has some kind of proprietary right in their database, or doesn't it imply that there's a liability on behalf of the other search engines that would scrape Reddit

Nick Garcia

Database right? In particular, triggers a whole other set of things that we have to think about, but especially in the European context, and in whether or not database rights are like their own thing, but in terms of their ownership of material. It's a really interesting question. And the very interesting thing about these licensing deals to me, and you know, so it's interesting bringing opening that can of worms a little bit, is that we don't really know what these licensing deals are about, like you laid out at the beginning. I think that this model of talking about data, collection, training and outputs is very important, because there's many different intervention points at which you might want to, like plug data in at different levels, or why you might need Intellectual Property rights or licenses at those three different steps, you might be getting a license to do the collection, or getting access to simply do the collection in a legal way. You might be trying to get a legal license to do training, because you think maybe you need that, I think you probably don't, but, or you might actually be looking at the output step and saying, We want to have inputs into this output step where we want to refine these results, or we want to be able to use specific components that might raise some copyright questions, and we want to have a license for those outputs. So there's all different reasons why these licensing deals might exist. And they're actually very vague right now, it would be a very important thing, I think, for us to find out what is actually going into these deals and why they are happening, what they're exactly for, what these companies are thinking about as they're exchanging them. Because all we know is that there is some deal going on for some amount of data changing hands?

Tim Lordan

Yeah, I've tried to find the ins and outs of these licensing deals. I don't know what they mean, but it's like the news The Wall Street Journal, News Corporation. Is it News Corporation, or is it, I forget what they call that company. Now that was like \$500 million. Do you spend \$500 million if you don't need to, Josh?

Josh Rogin

No, it's a really good question. And and I think the data question is really important, because you want these large you want good data, you want the results to be reputable and correct, and so you want some of this top tier data from reputable newspapers, is important to include, if you if you're only having older data, I think I saw that, you know, white men are responsible for most of the information on the Internet, so if we're using bad data, you're going to get bad results. So these are important questions, and. And kind of where it's like a plane has already taken off, and these models are already operating. And, you know, I tried to do a search in one of them a couple days ago for the political question, and they wouldn't let me, they wouldn't let me ask that question, which is just interesting. You know, we're still in the really beginning stages of this.

Tim Lordan

So, I wanted to go back to robots.txt, and that raises the questions of like, how can whether it be a website or a social media company or a video site or an audio site, how can they, are there anything they can do under existing law to say, don't ingest or train on my data. Is there anything they can do? Is robot.txt that? What happens with robots.txt, which says, "Don't scrape me, bro," basically? And does that? Is that just like a suggestion, or like, it would be nice? Or is there some, is there some kind of rule when it comes to how the Internet works, or how the law works, and is there something that Congress could do to make that rule more strong?

Brandon Butler

So I'll say, I'll start there's no copyright rule to that effect. And I would say that's a good thing. This is actually a really, really, really old issue. You know, back in the days of, you know, department stores selling, you know, dime novels. There was a case where the printer of the novel had put a little legend on the inside that said, "nobody can buy this for less than \$1" Right? And, you know, the department store sold it at a discount, and the publisher sued, and the court said, "Well, I'm sorry, once you have sold that copy to the department store, the department store gets to decide what to do with that copy," right? That's the First Sale Doctrine. And once you own a copy, the fact that there's magic words written on the inside isn't magic. It won't, it won't prevent you from doing what you want with your property, your copy that you own. And similarly, you know, when a web crawler encounters a website on the Internet, it's there, it's available, it is accessible, it's been made available. And there's no magic words you can put on the website that's going to trigger a new copyright obligation to prevent you from using that website. And I want to say that's a good thing, because without that, we wouldn't be able to do research right on the Internet. Researchers would be stopped from crawling websites in repressive countries, right, where there's all sorts of ways of suppressing information, but the copyright law doesn't, doesn't let you sort of opt out and say, "Actually, no, nobody gets to do Fair Use with this item," right? So that's, that's not the way that that copyright law works. There are other mechanisms, though there's contract

law, right? So, and this gets back to this question about deals. You know, why would you do a deal with someone like News Corp? Possession is nine tenths of the law, right? Like, if, if, if you have custody of your data, and no one can get to it. You can control it, and lots of data is in the hands of these folks. And literally, they can, you know, they they can provide the AI companies with more and better access than would be available on the open web.

Tim Lordan

Before I let Josh go, so you put up. Robots.txt, like, "don't scrape me, bro." No. Robots, scraping, anybody's free to just blow past that?

Brandon Butler

Under copyright law, yes,

Josh Rogin

no. And I think, which is why I said earlier, is that if copyright law isn't fit to purpose for this question, then Congress might respond so, so that's why this COPIED Act is not copy, copyright law. It is a new idea that, yeah, you can put provenance information and then you can determine who can use that under the law. So those are, you know, those are the what's coming if, if lawmakers aren't satisfied with how copyright law satisfies these questions.

Tim Lordan

Well, let me, let me just I'll let you go back. But that kind of, you're raising the question about, what is the purpose of copyright law, right? Like, why is there a Copyright which is kind of in the Constitution? Don't throw anything at me if you don't disagree with that, because this issue gets really heated. And the last time you know, Congress has really dealt with Internet copyright. And again, don't throw anything at me if you disagree with me, which, as Cam noted, was two years after the creation of the Congressional Internet Caucus, right the Digital Millennium Copyright Act. That was really the last time Congress has really addressed Internet Copyright. You could say the Music Modernization Act is a little bit maybe, but generally, that's that happens to be generally true. And the idea for copyright is in, again, I'm in the Judiciary Committee room, so if I misspeak, I'm in guerilla trouble, is to kind of incentivize people to create creative works. And if we're in this era of, you know, AI is training on massive data sets and of human generated content. What is the incentive going forward to, for humans to keep generating content? And what the 30,000 foot question is like, what impact writ large will have, this entire new space have on human creativity? People have been talking about like, "oh, it's super awesome, there'll be more human creativity." But ultimately, these data sets who will be contributing to them? That's the question.

Nick Garcia

Yeah. I love this question because I love talking about the super high level reasons for that.

Tim Lordan

We went from robots.txt to,

Nick Garcia

Yeah, And I think that something that's really important about this is, when you talk about copyright, it is excellent to focus on. It has two core purposes. It's that incentivization to create, giving people these limited rights for this period of time so that they can reap some return on their creative work. The other main purpose of copyright is to create a broad public domain that ultimately these works flow back into. That's the purpose of it being explicitly limited in the Constitution. Because what we want to do is incentivize people to create that way we can have this very broad shared corpus of human knowledge. That's what the Internet became very quickly, and I think we've seen that we don't need to do a tremendous amount to incentivize people to create. The real function that copyright has served, and that a lot of our economic systems have served, is we need to make sure that people have the material means to make a living so that they can create. People want to create. They want to create a tremendous amount of stuff. And a lot of the stuff that's created on the Internet is created for free and is amazing. And so this brings me back to some of the earlier stuff about, "why would you do this deal, what is the cost?" And also this point of great point that Josh brought up of inclusion in these data sets and these models, one of the big things that I'm worried about with potential copyright action here or with opt outs like something like a mandatory robots.txt might be, is that you could wind up in a situation where the only people who can then afford to participate in creating stuff or in contributing into these models are very, very big companies and very, very big media companies that are doing deals together to both generate content that is corporate, created content that they want to make, and then sell it to AI companies to put it in their models. And then the only people that are able to make AI models and do anything like this are the people that can afford to pay the exorbitant cost back to the media companies. This would be a very bad closed loop cycle for human creativity, for Artificial Intelligence, for media, and it goes against the core purpose of what copyright is supposed to be about.

Tim Lordan

Josh, you have any perspectives on Constitutional first principles on purpose of the copyright law?

Josh Rogin

Look, the, you know, copyright has been in this country, in our Constitution, it's, it's, yeah, it's the bargain we, you create, and then others are inspired by your creation and create anew and, so the question is, will these large language models break that, that deal? And that's what we need to we need to figure out, you know, will creativity continue? You know, part of humanity or or is this, this, these new giant tools going to interrupt that cycle?

Tim Lordan

Yeah, like I put up. So I was renovating a bathroom, and I learned how to do bathroom tile, not very well, but I put up a few Reddit posts about, you know, appropriate use of the bathroom tile over like a certain membrane, using a certain type of thinset mortar, maybe the three posts, and I thought they were just absolutely worthless, because they pretty much are. But if you take all those posts of people putting up worthless stuff on Reddit or other websites, taking it as a whole, right? All that, all those little bits of information about thinset mortar by a total amateur, taken as a whole, thrown into these large language models that becomes incredibly valuable. Hence the licensing deal, right? What? What incentivizes people like, Reddit looks like they made money. I assume whoever licensed that content is going to make money on the AI models. But what about me and my thin set mortar posts?

Brandon Butler

Are you going to stop?

Tim Lordan

Oh, well, more, when the tile falls off the wall, maybe I'll delete the post, but, but, I mean, that's the ideas like that. These little bits of information that we as humans contributed to the internet over time has been incred-, become incredibly valuable in this context, but there's no mechanism to compensate me for my creative work.

Brandon Butler

Well, I think interestingly, then it goes back again to this question of, what is copyright for, right? And copyright is for giving you a reason to do these things. And again, you didn't need, you don't need to be compensated to have this conversation, right? And. And the facts that would be derived from all of that raw information, you know, what's the best thin set mortar to use on this membrane? As we discover by reading 1000 amateur posts. Those belong to all of us. And we've, we've been down this road. You know, there were cases in the early 90s about telephone books and other forms of databases. And there was a movement at that moment when databases were taking off. And Nick alluded to sort of, there was a hot second where there were database laws sui generis, you know, non copyright, forms of protection for databases. And part of the rationale was exactly what you're describing. It takes work to make a database. Shouldn't we protect it so people can get the value? And in the United States, we had Case Law that said, "No, those facts are in the public domain." Now, if you can creatively select and arrange those facts, there's a layer of copyright on that, and you can sell the collection, but each individual fact is in the public domain. And in Europe, they they went a different way. They decided to have a database right, and they gave a special right to databases. And once there was a database put together, there was a way to sue people for taking the taking the data. And it flopped. That is, there have been two follow up reports. You know, did the database right work? Did Europe become the leading database provider in the world? No, they didn't. Had no impact whatsoever. Meanwhile, in the United States, after this decision about phone books, to the database industry collapse? No, we have a perfectly thriving database industry. So it's, it's, you know, we should be careful in thinking about whether to add a new layer of property control on top of information that could be really useful and reused by people in ways we can't even understand or predict.

Tim Lordan

Okay, so we talked about the collection phase of this AI process. We talked about the training phase of this AI process. We talked a lot about copyright as one of the three Intellectual Property rights. Let's kind of get to the output, which kind of implicates a little bit of name, image and likeness and trademark and things like that. So the outputs, like what gets generated, whether it's the response that you get from Chat GPT or the image you get from stable diffusion or dolly or, I forget what the music ones are called. What, so the Copyright Office has basically said, and you guys will can expand on this, is basically the output of one of these AI, large language models is not copyrightable because it isn't a human actually creating that content. It's the same kind of, the same concept as, like, if a monkey stole my camera and took a photo of itself. I couldn't claim copyright on that.

Nick Garcia

Actual case, Actually, it did happen. Literally. Happened

Tim Lordan

It literally. So they've kind of ruled on it. I mean, I don't want to spend too much time on it, but that's basically where it stands on outputs, right? As far as, yeah,

Josh Rogin

No, I think these are going to be interesting issues as we go forward. Because, you know, motion pictures, we're already using AI in our motion pictures, and we're being in, you know, Sound Editors are using it, special effects are using it, and we're certainly granted copyright on our material. So I think this is a, this is going to be an evolution. I think Congress has waded into this step, into this outputs issue. In the N, I, L step, there's a, there's a bill called the No Fakes Act, bipartisan, bicameral bill sponsored by Senator Coons and Blackburn, Tillis, among others, and that would essentially protect, give everyone a right of their in their own, a property right in their own digital replica. So this would protect singers from having their voices stolen, you know, actors from having their images taken without their compensation and credit and pay. And it would protect everyone you know, young women in high school that are having this terrible pornographic images circulated. So, so MPA has endorsed this bill. We think it's the right approach, and it has the right constitutional protections in it that allow us to keep making the the movies we're making, and at the same time, give people a right of over their own images as they come out. Yeah, it has a broad group of supporters in the in the artistic community, and then also OpenAI and IBM as well. So you know, on the output side, we think that's a really good way to protect people, give them a right over their own images and and accommodate for the Constitutional protection, yeah.

Tim Lordan

This is also similar to Scarlett Johansson's voice being used as an AI, you know, chatbot like her. Even though that that might not be a case, that's more of a we're not sure where that's going to go, but that's, that's the issue. Nick, Brandon, how do you feel about the No Fakes Act? Digital, we're calling them some digital replicas. Is that what we're calling them? How do we feel about that from a Constitutional perspective?

Nick Garcia

So just to go back first, super quick, to the copyrightability thing, in terms of whether or not these outputs should be protected or can be protected. I think we'll see some evolution on this thinking very soon. I think the best comparison for this is photography, which seemed like a very mechanical reproduction kind of a thing that people were very skeptical of how to do protection or consider protection, and it's arguably perhaps overprotected right now, but I think we might develop language about how to think about, is there some component of human authorship or contribution to AI works, especially as Josh was saying, when you're combining it in with a lot of other things.

Tim Lordan

And there's no, there's no Federal name, image, and likeness statute right?

Josh Rogin

There's, there's not, there's a lot of "Right of Publicity" statutes in the states, not every state, but most have it, and that's dealing with commercial purposes. You can't take someone's image and use it for commercial purposes that's already illegal.

Tim Lordan

And some of them have different, some of the different state statutes differentiate between image and audio and like it the different patchwork of different state rights of publicity and name, image likeness are all different, right?

Nick Garcia

Yeah. And on the name, image likeness and digital replica stuff where we're talking about deep fakes or reproductions of someone, that it's their voice model or it looks like them, I think we have a whole separate set of issues. The way that we are thinking about this right now is we have three sets of potential harms that we have to really worry about. You have potential economic harms in terms of people who want to be able to make commercial use of their likeness in some way, want to protect the commercial value of that by licensing it out, for, to appear in a commercial, or something like that. Or you don't want to be seen in an ad endorsing something that you don't like because you're a famous person, you have issues related to our human dignity, in terms of, as Josh brought up, some of this horrible stuff in terms of non consensual, intimate imagery, or ways that people use these technologies to harass or abuse people, and especially women. And then we have a third category of things, which is potential democratic harms, in terms of, are these deep fakes? Are these voice models being potentially used to create and disseminate disinformation, things that are going to mislead people. I think we need to think about all three of those buckets of harms together when we're looking at any kind of a Legislative fix for it. I think the No Fakes Act is something that overreaches in some ways by creating this very new broad Intellectual Property right, and it doesn't really properly balance between all three of those things, very briefly, I think one of the biggest issues there is that it creates a Notice and Takedown regime, as you said, very similar to the Digital Millennium Copyright Act, which was the last time Congress touched anything related to the Internet and IP. And that takedown right is super overbroad right now, and it's very difficult to get anything put back up or to have stringent requirements about what is getting taken down. And as a result, I think it really fails in that democracy category that this could become a massive tool for censorship, is the takedowns under something like No Fakes. So that needs work that has to be tuned in order to make sure that we're getting this right. Brandon wants to say something, you guys. I'll go to you guys in one minute for questions, so get your questions ready. Brandon. And, that's part of why state name, image, and likeness rights are not Property Rights. They're a separate, they're usually contracts or torts related things, so

Brandon Butler

Very quickly, there is a, there are at least two serious Constitutional problems with No Fakes. The obvious one is the First Amendment, right. There's a free speech issue. We have, we have a free speech First Amendment Right to talk about other people, to, you know, to use images in a creative way. We've done that for 150 years. People, you know, have been caricaturing people, using photographs of people in new and interesting ways. So we have to think about, it's not the case that, you know, there should be no regulation of digital replicas, but we do have to think about how that regulation impacts the rights of others to use them in a creative way. So that's a one Constitutional

question. The other is the Constitution tells us what kinds of Intellectual Property Rights Congress can create, and it says "Congress shall have the power to create 'exclusive rights' in the writings of authors." And the court has said that means creative works, again, not factual information, so to create a Property Right in the fact of you know, how I look and how I sound is arguably, Constitutionally not permitted. It's not something Congress can do. So they'd have to find a different way to go about these problems. And I think those other ways would have the advantage of being more targeted in the way Nick describes, they would find the harms and focus on those, rather than creating a Property Right. All the Property Right does is say you can buy and sell something. It doesn't necessarily protect anybody.

Josh Rogin

And if I could just respond quickly, I think. It would be news to Courts and Congress that they don't have that right. I think trademark law, trade secrets have been eminently Constitutional. And I would also say that you know this, if you want to take down right for a non-consensual deep fake, if California has a law in the books, which they do. It's not able to have it taken down from social media that's shown all over the world. This is actually an issue that you need Congress to act on. So we have uniform rights. And I just from the you know, we at the MPA, we care deeply about the First Amendment. Our business is based on the First Amendment, so we were very careful with the Constitutional protections this bill and the ban does not include when the replicas are used consistent with the public interest and bona fide commentary, criticism, scholarship, satire or parody. So that gives you an idea of how broad those exemptions are, and they can protect the speech like satire and parody that you're mentioning.

Tim Lordan

We're just we're just touching on these issues. And Josh earlier mentioned that some huge percentage of the Internet content is created by white men, which is not good. And it goes without saying that this is a panel of four men, and we could spend an entire time focus on, focusing on deepfakes of sexual imagery of women, and that that probably will be an entirely separate panel, because there are so many issues there. And I'm not trying to not give it, do it justice, but we just don't have time. I think we're going to save it for another briefing next month, or the month after we did something very similar years ago on when it comes to non-consensual imagery when stolen, whether that's a copyright issue or whether that's a tort, etc. So we'll get we'll get to that. But let me, let me go to questions right there.

[Audience question inadubile]

Josh Rogin

Yeah, those are interesting lawsuits, but there's a automatic right to to play music, and there's not, you know, in Europe, they have some. They have, like, moral rights sometimes that go along with their Property Rights. But in this country, unfortunately, you can you get an automatic instance to play a song no matter who's playing it. So they are being paid. But there's a reason they're sending cease and desist letters instead of lawsuits.

Nick Garcia

I think it's a good example of how copyright can serve as a really big hammer, an attempting one that people want to go lean on using Intellectual Property and copyright when they see something going on

that they don't like in the legal system. And I think we're seeing that right now with the AI. Goes to what Josh was saying about is copyright fit for purpose here? But I think that cuts both ways. I think the question is, is copyright the right tool for addressing some of these problems that we see with AI, I think probably not, is the answer?

Tim Lordan

Okay, right here

Audience question 2

Thank you. My question is about the commercial reuse of AI, voices and image likeness. So what if you're born and look like or a sound alike? Are you prevented from commercial sale of your name, image, likeness voice, to AI organizations that then it gets reused commercially? If someone looks exactly like sounds exactly like Scarlett Johansson, so naturally, are they prevented from ever becoming a commercial actor or selling their likeness that then gets used commercially in any number of those organizations. What, what threshold are they then prevented it from being because there's now a personal statement or doppelgangers going, Hey, I don't want you to sell my image and likeness, because you look just like me.

Tim Lordan

And that is similar to the Scarlett Johansen OpenAI "Her" chatbot, where the voice they actually use did sound a lot like her, but maybe wasn't.

Josh Rogin

Yeah, no, these are, these are excellent questions. And yeah, the Scarlett Johansson case is interesting because they, yeah, they, they hired someone separate that had a similar sounding voice, but it was not Scarlett Johansson. And I think there's been earlier cases with Tom Waits, with a sound alike, also where you can, you can't prevent, you can't prevent, like, a style of of singing to, but, but those, these are going to be interesting questions as we go forward. And, you know, yeah.

Brandon Butler

Yeah, and I think what your question points to is, again, the need to balance different interests, right? I mean, you're exactly right. There are people who. Want to act and participate in entertainment culture who have the voices they have right, and we need to balance their interest in having a career with the interests of people who came before them, who in protecting their reputation, for example, and and there are aspects of the way name, image and likeness rights are enforced that helped to do that. So if no one would believe that was Scarlett Johansson, for example, or, you know, then most name, image, likeness, rights regimes are not going to punish that person, so the person who did that VoiceOver is not in any danger. But, you know, a new Property Right, the thing about property rights is they are. It doesn't matter what your state of mind is. It doesn't matter if anyone was harmed by what happened if you violate a property right? End of story. And we don't have room to say, well, was anybody harmed? What are the other interests here? So that's one reason we're worried about a property based approach, like No Fakes, rather than a harm based approach like we've seen in the past

Tim Lordan

In the second row, Adam,

Audience question 3

Yeah, I'm Adam Eisgrau Chamber of Progress. We're a center left trade organization for technology companies that can fall including generative AI producers. I want to take issue with some of my colleagues who are actually on the same side of the, disparity, since ever an AI issue, we get them to comment on that, take issue briefly with Josh. The first point is that the Constitution absolutely is relevant here, that's where copyright comes from. Nick, I believe, said that there are two purposes of copyright, to promote progress in science, abuse law and to pay people to incentive them. My issue is maybe equivalent, but it's important to understand the purpose is to promote progress in science and useful arts. It's not sacrilegious. That's the means to me.

Tim Lordan

Totally but if you could matriculate to a question, The question is, is that right, or am I simply [inaudible], which I am

Nick Garcia

yeah, no, the That's right, and I didn't mean to imply otherwise. The purpose of copyright is to promote that progress is to promote, as I said, the creation of that big, broad public domain, and we do that by securing to them those limited rights for limited period of time in order to incentivize them to create, which is the other purpose, is that incentivization. And I think that goes back to some of what we were talking about earlier, in terms of what's then the incentive to make the Reddit post right? Like, what's the incentive to continue making creative work in this space. And I think, as I said at the very top, we probably want to think about different things for different kinds of works. The incentives creating a big Hollywood movie are totally different than the incentives for creating an informational Reddit post. And so I think we should disaggregate out the idea that the only way to do this is through compensation. One thing that we might want to think about are, what are the benefits of these models that are being built? And there's gonna be different kinds of models doing different things. I think we wanna think about, what are the public benefits of AI, and is this something that is serving us as a society? There are interesting ideas that we are working on around the concept of Public AI. We could think about how to implement requirements on AI companies for them to be returning public benefit in the same way that we have a long history of doing that for media companies that were required to have public interest kind of benefits that they would have to return in exchange for things like their broadcast licenses. There's many ways that we can think about like, how do we actually continue to make sure that these things are beneficial so that people feel good about continuing to both create work and most importantly, I think, share it on the Internet for other people to use.

Tim Lordan

I think a question in front row right here

[inaudible] my father was part of a band called The Commodores,

*clapping

We have a an IP case where we see a lot of the work that he's created now is being used and he's not getting credit for it. My question for is, How can we create policies that protect entertainers, influencers, and creators?

Josh Rogin

Yeah, I think, I mean, that's what's what's motivating a lot of the No Fakes Act is to give people control over their images. And then, as we were saying, These are the questions that we're, they were considering, like with the with the COPIED Act, which I keep mentioning just because it's an interesting counterpoint, which is that would give people more control over their their their music, in your case, and that wouldn't be able to be used within some big AI music making machine. So these are the exact issues we're talking about, and whether you pull out things out of the data set or not, these are issues that Congress has got to decide. And I think it's worth actually, while we're winding this up is, is Congress capable of doing the big things? Is Congress even capable of legislating on AI? And you said 1998 was the DMCA, you know, I was, I spent 20 years up here on Congress, and try to think of the last time that both parties appointed representatives to work on an issue together, and that's actually happening. You know, in the Senate, Schumer appointed a bipartisan group of senators. In the House, that happened as well. And I don't, I don't know if I see a radioactive partisan issue at the center of all this that can spoil it all. So I think it's possible that Congress can do this, but it's going to be hard. It's hard for Congress to pay the bills. So as we're thinking about these court cases and Congressional response, it's worth asking is any can this body can do the hard work that it needs to be done in the AI and IP space?

Brandon Butler

I want to add one, one more thing, and it also goes back to the Isaac, and

Tim Lordan

That better be put the Commodores on your playlist

Brandon Butler

After you put the Commodores on your playlist, the ways that musicians are often harmed right in the music business is that their rights come out of their hands very quickly and are dealt with in ways that they don't approve of. Right, creating a property right as No Fakes does that's licensable and saleable does not necessarily mean, then, that the people who get that right at first will be well served in the long run. So the disputes like the ones you're describing, it's often the case, right, that a manager, a record label, somebody down the line, sort of fair and square got a hold of the rights at issue. And the debate now becomes is, was that fair and square? Who holds the rights? Can I stop them? And it's not necessarily the case that the existence of a property right is going to answer that question the way that we think it should. And similarly, for the Isaac Hayes question, I think part of what's happening in that litigation is the question is, you know, did Isaac Hayes agree to have his songs played in that context or not? And as historically, it's exactly as Josh described. Typically, artists have not had much control over that, but in the last few years because of exactly what we're seeing, what what Mr. Hayes is litigating about, they've taken some of that control back, and they've put clauses in contracts that say, I want approval for political uses. But that's all about power between the artist and the other people in this

ecosystem. And property doesn't necessarily create power. It just creates a thing that flows, right, in accordance with the power dynamics that already exist.

Josh Rogin

And I would just say in the No Fakes Act, we were very careful about, the, Coon's office is very careful about the licensing provisions. And I think gives right and power to the licensees so they, you know, you cannot sell your your digital replica for 50 years when you're a teenager. So, but I so I just want to make that point.

Tim Lordan

I think there's the Recording Industry, so Recording Industry Association of America has brought a lawsuit against two audio AI companies, and they've actually gone through the first part of like discovery and presentation of facts. There are a lot of facts that'll be really interesting in that as that case plays out, but it's just started, but that goes directly to the audio component of that's only one part of your question, but that'll play out for the next year and a half or so. There's so much I didn't get to in this panel. I am so sorry, but I did promise we'd only be here for an hour. I want to thank the panel. I want to thank you for coming. Want to thank the Congressional Internet Caucus for hosting, and thanks so much. Appreciate it.