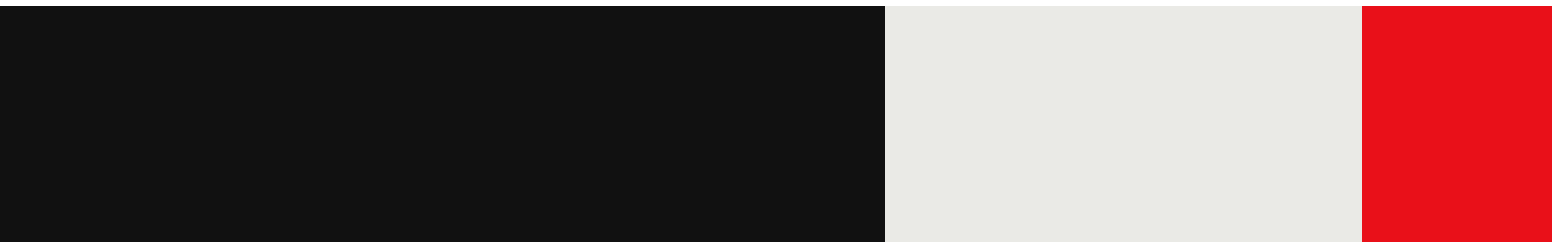


The Daily: Is copyright genAI's kryptonite, how 'fair use' comes into play, and what happens to the creative arts?

Audio



On today's podcast episode, we discuss how copyright lawsuits could down OpenAI (or the whole industry), whether publishers will land on The New York Times side of the generative AI (genAI) copyright debate or on the Axel Springer and Associated Press side, and how copyright will impact the creative arts. Tune in to the discussion with our analysts Evelyn Mitchell-Wolf and Yory Wurmser.

Subscribe to the “Behind the Numbers” podcast on [Apple Podcasts](#), [Spotify](#), [Pandora](#), [Stitcher](#), [YouTube](#), Podbean or wherever you listen to podcasts. [Follow us on Instagram](#)



Episode Transcript:

Marcus Johnson:

This episode is made possible by Nielsen. What even is TV anymore? I don't know. I guess it's phones, tablets, refrigerators perhaps. So, make sure you are looking at everything. Check out

Nielsen's Upfronts/Newfronts guide for the full picture. You can get that at nielsen.com. It's right on the homepage.

Evelyn Mitchell-Wolf:

Funneling people away from The New York Times websites, which makes ad revenue harder to generate. Then, why would people subscribe to The New York Times if they can access most of the information that The New York Times puts out through ChatGPT? It is cannibalizing those revenues.

Marcus Johnson:

Hey, gang. It's Tuesday, March 19th. Evelyn, Yory, and listeners, welcome to the Behind the Numbers daily, an eMarketer podcast made possible by Nielsen. I'm Marcus. Today I'm joined by two people. We start with our senior analyst who covers everything digital advertising and media. Based in Virginia, it's Evelyn Mitchell-Wolf.

Evelyn Mitchell-Wolf:

Hi, Marcus. Hello, everyone.

Marcus Johnson:

Hello, there. We also have our principal analyst who covers everything advertising, media, and tech. Based in New Jersey, it's of course Yory Wurmser.

Yory Wurmser:

Hey, Marcus.

Marcus Johnson:

Hey, fella. It's another morning recording, so I'm going to complain about how difficult mornings are. I got up the other day for a 6:30 yoga class, believe it or not, and I was so proud of myself. I was like, "Yes." I only went because someone made me. I got there and it was rough. But then afterwards, it was 7:30, I said, "Maybe I'll get some coffee. Maybe I'll read." I went right back to bed for three hours.

Evelyn Mitchell-Wolf:

Oh, no. Marcus, you did so good.

Marcus Johnson:

It would such a failed attempt to be a good person.

Evelyn Mitchell-Wolf:

I find yoga in the morning is so much harder because you stretch out throughout the day, and then [inaudible 00:01:56] would-

Marcus Johnson:

Thank you.

Evelyn Mitchell-Wolf:

After work, you're already-

Marcus Johnson:

That's what I said to the yoga teacher. I said, "Why is this even happening?" She didn't care for, kicked me out. Anyway, today's fact, Atlas is holding what on his shoulders? Do you know-

Evelyn Mitchell-Wolf:

The world?

Marcus Johnson:

Greek myth-

Yory Wurmser:

The world, yeah.

Marcus Johnson:

No, we've been lied to. Well, he didn't. I'm sure he didn't start the lie. "What's that on your shoulders?"

Evelyn Mitchell-Wolf:

The sky.

Marcus Johnson:

"The world" Basically, yeah. Yeah, the heavens. Atlas is not holding the world on his shoulders. In Greek mythology, the Titan Atlas was responsible for bearing the weight of the heavens on his shoulders, a burden given to him as punishment by Zeus after the Titans revolted against the Olympians. There are a lot of pictures of him holding the world, but there are some of him holding just a globe-shaped object.

Evelyn Mitchell-Wolf:

Yeah, 'cause how do you depict the heavens?

Marcus Johnson:

Yeah. There's different symbols on one. They all look a bit different. But, he's not holding the world.

Yory Wurmser:

I blame Rockefeller Center for this. They have a statue of him, and he has the world or a globe on it.

Marcus Johnson:

He does?

Yory Wurmser:

I think so.

Marcus Johnson:

Liars. That's a tough sentence though, isn't it? "Here, hold the heavens forever." Atlas is like, "Can I take the jail time, or at least appeal?" That's rough. But yeah, he's not holding it. The heavens does do a lot. It's a fair play. Anyway, today's real topic is generative AI's Kryptonite's copyright. In today's episode, In the Lead, we'll cover whether copyright law can stifle GenAI's world takeover. No In Other News today, there's too much to get to. We start, of course, with the leads and this generation's big copyright battle pitting journalists against AI software that has learned from, and can regurgitate their reporting rights.

Adam Clark Estes of Vox, in his piece, he references P2P file sharing site Napster, which made it easy to download music for free before the days of Spotify and Apple Music. Record companies were quick to say, "Not so fast," and in 2001, a federal court ruled that Napster

was liable for copyright infringement. In a modern version of this case, The New York Times is claiming that OpenAI trained its model with copyrighted Times content, did not pay proper licensing fees. Mr. Estes says, "The consensus of casual observers and legal experts is that this New York Times lawsuit is a big deal, since not only does The Times appear to have a solid case, but OpenAI has a lot to lose, perhaps its very existence." So Evelyn, I'll start with you. What are your thoughts on the idea of copyright lawsuits downing OpenAI?

Evelyn Mitchell-Wolf:

I think they could, but I don't think they will. I think it's more likely that copyright lawsuits will reshape what a viable business model looks like for GenAI companies, including OpenAI. The initial business model started with scraping every available piece of content from the internet, regardless of whether there were permissions in place or not. Then the next steps were to use that data to train the model or models, put the product on the market for free, foot the bill for the computing power, get potential paying clients hooked, then charge for access. That's where we are now, and there are a lot of ways that the resulting subscription model can look.

I don't think it necessarily all falls apart if OpenAI, or any company decides to or becomes legally obligated to pay creators, publishers for the content they use to train their models. Training data just becomes a larger component of open AI's overhead, and then the cost of operating a GenAI business goes up. Then that cost increase gets passed along to end users the same way that the price of takeout goes up when the price of ingredients goes up. How much profit margin OpenAI will be willing to sacrifice under those circumstances to keep costs palatable for its subscribers, that remains to be seen.

I know it's not as cut-and-dried as I'm making it sound, of course. Training data isn't just used in perpetuity, it's ingested into the model. GPT-4 can't unlearn what it has learned so far by training on the unlicensed New York Times data. So this gets into the whole transformative conundrum of the fair use doctrine. I just think the outcome of these lawsuits will reflect both the pervasive ideology of innovation for innovation's sake, and the fact that GenAI is an existential threat to the publishing industry as we know it today.

Marcus Johnson:

You touched on a lot of interesting points. One of them is the question of, how much is OpenAI willing to pay news outlets? There is a report in the information from January saying, "OpenAI offered some media companies as little as between \$1 to \$5 million to license their articles for

use in training its large language models." Yory, when we were prepping for this episode, you were saying, "Big question is, how do you make money whilst paying those very copyright fees?"

Yory Wurmser:

Yeah. I just want to agree with what Evelyn says. I don't think it's existential for the AI industry, but I think it's going to be existential for some AI companies, just because the business model will change with these higher fees. I think some of these suits are going to... Well, I'm not a legal analyst. But to me, under fair use, it seems like a lot of these companies are going to have to pay higher figures for some of this data, and that's going to make the viability of some companies a little harder to sustain, and probably raised prices for AI services. So my guess is, the million dollars, or \$5 million, or whatever they're offering a company like New York Times, it's just not going to cut it.

Marcus Johnson:

Although they probably want to pay for permission then pay for forgiveness because some of these fines could be company-ending fines. It depends on the size of the company, you're right. But The Times is arguing, "OpenAI is making money off of content and costing the paper billions of dollars in statutory and actual damages. By one estimate given the millions of articles potentially implicated, and the cost per instance of copying, The New York Times might be looking for \$450 billion in damages," as Mr. Estes of Vox. You mentioned fair use theory as a huge part of this. Just to set the table on fair use, we'll talk about it for a second.

According to the Copyright Alliance, they say, "Fair use permits a party to use a copyrighted work without the copyright owner's permission for purposes such as criticism, comments, news reporting, teaching, scholarship, or research." "However," as the University of Illinois points out, "If the purpose on the other hand is to make a profit, or for commercial gain, that would weigh against fair use." Nilay Patel of The Verge explaining that, "Since the law can't predict what everyone might want to do, there's apparently a four factor test written into it that courts can use to determine if copy is fair use." Any courts get to run the test any way they want, and one court's fair use determination isn't actually precedent for the next courts," you'll be pleased to know. Evelyn, when it comes to fair use, anything's possible in terms of what could happen as an outcome.

Evelyn Mitchell-Wolf:

Even just the idea that precedent is not necessarily established when one court goes through the whole exercise of looking at the facts of the case and determining whether fair use applies, that makes this whole can of worms even more complicated here. I think that commercial side of the equation, like you mentioned, it really does weigh in The New York Times favor here. When we think about, also, there's another element of fair use. Which is, the product that is the result of the copying, is it replacing the original product in terms of market share and the ability of the original producer of the content to continue their business, and continue to make money?

That is also not so much in OpenAI's favor because even if you consider, the training data aside, there's also this existential question of LLMs, ChatGPT, whatever have you, funneling people away from The New York Times and other publishers websites, which makes ad revenue harder to generate. Then, why would people subscribe to The New York Times if they can access most of the information that The New York Times puts out through ChatGPT? It is cannibalizing those revenues.

Yory Wurmser:

There is one more element of fair use, which is just public information. So if The New York Times is reporting on some public news, they have a much harder case. But if the analysis is being reproduced, then they have a much stronger case. But I think the impact on the publishing industry is that the long tail of publishers is really going to be hit. They're not going to get traffic from searches nearly as much because the answers are going to be there for this public information, and it's going to just be a lot harder for them to have that differentiated, non-public information that they could sell and defend against.

Marcus Johnson:

There's this question of, what if regurgitation is eliminated? This is when OpenAI generates text that matches Times articles word for word. The Times provides lots of examples in the lawsuit against OpenAI, but OpenAI says regurgitation is a "rare bug" that they are "working to drive to zero." So I don't know if that's going to change any judges minds. OpenAI's solution to this battle with The Times is to pay the copyright owners first, as we talked about, the way they've struck licensing deals with folks like the Associated Press, and our parent company Axel Springer. But Yory, do you think the industry's likely, at this point, if you had to speculate, land on the side of The New York Times in terms of the GenAI copyright debate, or on the Axel Springer and Associated Press side?

Yory Wurmser:

I think the industry is going to diverge here. So the big publishers that can charge subscriptions, I think they're going to try to get a good deal as they can, or sue OpenAI. I think the smaller businesses, the small publishers are screwed because they don't have the resources to fight nearly as easily. I think their strategy is more going to be AI optimization, try to be one of those links that the AI bots then link out to when they create their single answer, and possibly join industry groups to have collective payments or something. But in terms of The New York Times strategy of driving a hard bargain, I think that's going to be limited to a few large publishers, and the bigger publishing landscape to long tail is going to have a much harder time doing that.

Evelyn Mitchell-Wolf:

I also looked at it as, the industry was going to diverge, and I segmented the divergence among publishers versus advertisers and ad intermediaries, because publishers are going to feel very differently than advertisers and intermediaries, obviously. Like you just mentioned, Yory, they have different stakes in this. I think though that most publishers want The New York Times to win because if The New York Times wins, all publishers win. There is that matter of, precedent doesn't hold here, but very few publishers have the financial resources and the brand recognition to fight this battle in court.

If The New York Times wins, that would establish some... It's not precedent, but something in court, some reference material for publishers to point to and say, "Okay. Well, we know how this went this time. Do you really want to gamble, OpenAI? Do you really want to gamble, Google, on going to court again to ask for forgiveness rather than permission?" So The New York Times, if it wins, that's a big deal for every publisher. When it comes to the way that Axel Springer or the AP have approached partnerships with OpenAI, we all talked about that, that the compensation is probably not enough to recoup what they're going to lose in ad and subscription revenue in the long run.

It's better than nothing obviously, but it is only for those large publishers, like you mentioned, Yory. There are those tons more publishers who have no shot at striking those kinds of licensing deals. So if New York Times wins, there's going to be a positive knock-on effect for the whole publishing industry. One of the points that Nilay Patel made, that I thought was a really good one, is that if it's not The New York Times now, if The New York Times loses, other publishers will come in to fill this lawsuit void, because it is such a huge deal to so many

publishers through Authors Guild. If we have time to talk about the music industry, their artists are all in on this. So I can see this playing out in courts regardless of whether this one case goes in The New York Times favor.

Marcus Johnson:

One final point before we go to the creative artist side of this argument. There was an Atlantic article by Alex Reisner citing William Patry, a former senior official at the US Copyright Office, saying a blanket ruling about AI training is unlikely. Instead of saying AI training is fair use, judges might decide that it's fair to train certain AI products but not others, depending on what features a product has or how often it quotes from its training data. We could also end up with different rules for commercial versus non-commercial AI systems as well.

But Evelyn, to your point about the artist side of the argument, Mr. Estes at Vox was explaining that, "The New York Times are not the only party suing OpenAI and other tech companies over copyright infringement." There's a growing list of authors and entertainers that have been filing lawsuits since ChatGBT hit the scene, accusing these AI companies of copying their works to train their models. AI companies have argued their language models learn from books, and produce transformative original work just like humans. Developers are also suing OpenAI and Microsoft for allegedly stealing software code. Getty Images is suing Stability AI, the makers of image generating model stable diffusion over its copyrighted photos. Yory, how do you think AI copyright's going to affect the creative arts?

Yory Wurmser:

It's a very similar argument to what we were just talking about for publishers. I do think that the creative arts can argue distinctiveness and non-public use a little more clearly than news stories, for instance. So, I think there's a strong case there. The challenge is, the artists will have to band together, unless you're Sarah Silverman or someone, who is suing. There will have to be some group class action suit that has resources to fight this.

Evelyn Mitchell-Wolf:

Absolutely. To your point, Yory, facts cannot be copyrighted. So there is a difference here between news stories and the more creative... We read an article where Hosier was quoted as saying, "AI can't really qualify as art because it's missing that human component." I think that element here is really interesting. If there were to be some collective action, like you

mentioned, is a key difference between The New York Times suing OpenAI and a group of artists suing Stable Diffusion or whatever, because there's this component of the human experience that a judge may be more sympathetic to. It's the risk of our very humanity here. It's intense and I think it is at stake. Artists are a really big part of what it means to be human. If the financial incentive to create art is flipped on its head, if artists cannot sustain themselves by doing their craft, what does that world look like?

Marcus Johnson:

It seems like plagiarism is certainly a problem. There is a new report from plagiarism detector Copyleaks that found 60% of open AI's GPT-3.5 outputs contained some form of plagiarism notes, Megan Morrone of Axios. We saw the Hollywood writers striking in part over the potential for AI to take over their jobs. Maybe its musicians turn next. I know Hosier in that article was such as saying that he would go on strike, stop making music, or some other form of strike. Another interesting part of this as well, a US district court judge recently ruling AI generated artwork cannot be protected under copyright law in a major legal decision saying that human authorship is a bedrock requirement of copyright. So, that being another component here. What happens if you make a piece-

Evelyn Mitchell-Wolf:

Let's just make things more complicated. Why not?

Marcus Johnson:

With GenAI. I know. That's what we've got time for, for this episode. Thank you so much to my guests for helping me pick through this incredibly complex issue. Thank you to, Evelyn.

Evelyn Mitchell-Wolf:

Thanks, Marcus.

Marcus Johnson:

Thank you to, Yory.

Yory Wurmser:

Always a pleasure.

Marcus Johnson:

Thank you to Victoria who edits the show, James, Stewart, and Sophie, the rest of the podcast crew. Thanks to everyone for listening in to the Behind the Numbers daily, and eMarketer podcast made possible by Nielsen. Tune in tomorrow to hang out with Sarah Libo as she speaks with our chief content officer Zia Daniell Wigder and principal analyst Jasmine Emberg, as they come to us live from Shoptalk in Las Vegas.