


The Daily: News and AI divided—the NYTimes sues OpenAI, other large publishers strike deals, and how this all shakes out

Audio



On today's podcast episode, we discuss the argument for the NYTimes suing OpenAI, the argument for other publishers (like Vox, the Associated Press, and others) signing content licensing deals with OpenAI, and how this all shakes out. Tune in to the discussion with our analysts Daniel Konstantinovic and Gadjo Sevilla.

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 (00:00):

This episode is made possible by Roundel, partner with roundel and reach over 165 million people who look to target for joy and inspiration. Together with them, you can design curated media solutions that are a seamless extension of the target experience, all backed by

unparalleled first party data and measurements. If you'd like to learn more, you can head to roundel.com.

Gadjo Sevilla (00:24):

What's important to point out though is that quality content is a finite resource, and so at some point, all the bases will be covered, right? And the way these training models just suck in all this information, this is not going to go on forever clearly.

Marcus Johnson (00:46):

Hey gang, it's Thursday, June 6th. Danny Gadjo and listeners, welcome to the Behind the Numbers Daily and eMarketer podcast. Made possible by Roundel. I'm Marcus. Today I'm joined by two folks. First of all, we have one of our senior analysts who writes for our connectivity and tech briefing is Gadjo Sevilla.

Gadjo Sevilla (01:05):

Hey Marcus, happy to be here.

Marcus Johnson (01:06):

Hey, fella. He's based in New York City, also based in the city is analyst for our marketing and advertising briefing. It's Daniel Konstantinovic.

Daniel Konstantinovic (01:16):

Hello. Happy to be here.

Marcus Johnson (01:17):

Hey, chap. So today we're talking all about AI companies and publishers and the divide that's going on there, but we start with today's fact. The deadliest animal for humans is, I'm guesses,

Daniel Konstantinovic (01:32):

I don't know mosquitoes. Is that a? Does that count?

Marcus Johnson (01:34):

Oh really, Danny, according to BBC Science Focus, they kill 725,000 people every year because they carry various deadly diseases, dengue, yellow fever and malaria, things like that.

They kill three times more humans than second place snakes. Mosquitoes are the worst people. They're so awful. Today's real topic, AI companies and publishers. How does this relationship play out?

Marcus Johnson (02:09):

In today's episode, we're just talking about the divide between AI companies and publishers. No, in other news. Okay. So we start with the lead. The news industry is divided over AI, right? Sarah Fisher of Axios explaining that major news outlets are taking opposite approaches towards future proofing their businesses against the threat of AI. Some are opting to partner with AI companies, others are suing them. On the one hand, the New York Times, some smaller news outlets and eight prominent regional players like the Chicago Tribune are all suing ChatGPT, parent OpenAI, and Microsoft for copyright infringement, basically using its data articles and content libraries to train its AI models. The New York Times had been in talks with OpenAI on a licensing deal for context, but that fell through sending them down this path of suing them. On the other hand, several large news publishers like the Associated Press, Financial Times, Meredith, Vox, the Atlantic, and others have instead chosen to strike pay deals with AI companies for millions of dollars annually, undermining the Times argument that New York Times that is that it should receive damage payments of billions of dollars.

Marcus Johnson (03:16):

Ms. Fisher notes. So she also points out though that AI firms don't need to license news content from every publisher depending on the type and the volume of content they need to train or inform their large language models. So for this episode, Danny has offered to be arguing for working against AI companies. So that's the New York Times position suing them. And Gago has kindly agreed to argue for working with AI companies, which is what Vox News Corp, Wall Street Journal, the Associated Press and others are doing. These arguments don't necessarily reflect these gents personal opinions. They're very kindly agreed to just be lawyers for the different arguments. So you can hear two different sides. So we'll start with Danny and we'll start with an opening statement from Danny. He's arguing for working against AI companies like the New York Times is.

Daniel Konstantinovic (04:04):

Sure. I do feel that on some level by signing these deals, publications are kind of giving AI companies the shovel with which to bury them. I understand that there is a big cash influx

from these deals that is likely really helpful to these struggling publishers. It's a really tough time for the digital publishing industry, but the products that OpenAI and Google are developing are clearly intended to replace these publishers or at least siphon off a significant portion of their traffic. If you just look at what Google's AI overviews are doing, that is a significant risk to publisher traffic. And I think that if publishers had successfully banded together and as there were talks to do less summer, according to the Wall Street Journal and sued AI companies for copyright infringement, they would've had a much stronger case than they do now with all these deals. But their position is undermined by their industry colleagues signing these deals.

Marcus Johnson (05:05):

Right. Ga, you're arguing that working with AI companies and signing licensing deals is a smart option. How come?

Gadjo Sevilla (05:14):

So I think they want to take control of the situation and they probably feel if you can't beat them, join them. So media companies working with AI companies to make their content available for AI data training or doing two things, they're finding new ways to monetize their content while trying to remain relevant in the shift to AI by making their information available to a new audience, they also have the opportunity to shape the future. A lot of these deals are not just for content, but they're also for building new content products with the oversight of the content companies themselves.

Marcus Johnson (05:55):

So Gaja, the pushback here, which is New York Times is concerned about, I mean you said if you can't beat them, join them. And some folks see this as kind of a deal with the proverbial devil because Jessica Lessen of The Atlantic, ironically, one of the companies who now has a licensing deal with open ai, she wrote that media companies are making a huge mistake with ai. News organizations rushing to absolve ai, companies of theft are acting against their own interests. It's a really interesting piece. She writes that for as long as I've reported on the internet companies, I've watched news leaders try to bend their businesses to the will of Apple, Google Meta and others chasing text distribution and cash use firms strike deals to try to ride out the next digital wave. They make concessions to platforms that attempt to take all

of the audience and trust that great journalism attracts without ever having to do the complicated and expensive work of journalism itself. And it never ever works as plans.

Daniel Konstantinovic (06:51):

You call it a deal with the devil. A day after they announced the deal, a columnist, the Atlantic published a story about the deal with that exact title.

Marcus Johnson (07:00):

Oh, really? I

Daniel Konstantinovic (07:01):

Think, yeah, lesson's also the founder and editor in chief of the Information, which is another tech publication that has significant stakes in what goes on with ai. And I think she's right. I mean she makes some really compelling arguments about historical examples of news outlets tying themselves to tech companies, whether that's the infamous Facebook video debacle where a bunch of companies infamously pivoted to video only to find that their Facebook's metrics were significantly inflated and a lot of companies went out of business. One of the examples I believe she brings up in the article is when NewsCorp developed a publication specifically for the iPad that burned through tens of millions and shut down less than a year later. Yeah, I mean it's hard to argue with all of the historical examples of companies being media companies being hung out to dry in these deals or partnerships or focuses on tech platforms.

Marcus Johnson (08:01):

I mean, you said if they can't beat them, join them and they're getting some money out of this, but it doesn't seem like that much money at all. And in the piece she talks about publishers, the idea of them willing to roll over this way and just failing to defend their intellectual property, saying they're also trading their hard earned credibility for little cash from the companies that are simultaneously undervaluing them and building products quite clearly intended to replace them. So what do you make of how much money these companies are getting in some of these licensing deals?

Gadjo Sevilla (08:34):

Yeah, from my understanding, most of these deals are non-exclusive, meaning a content provider can work with several AI companies at the same time, and so that just makes it possible for them to earn recurring revenue while ensuring that their content and coverage remains a staple for the training models models. This also helps, I mean, it's inevitable that the audience might be using AI rather than the web or search to get through articles. So I think they're just making sure that they're in that conversation. The content is still accessible through that means, I mean, it's not a perfect deal. Clearly they're giving up a lot, but at the same time, these affiliate programs can provide better returns than advertising for them.

Marcus Johnson (09:29):

It does feel like a bit of a cash grab and not that much cash at that. I mean, you make a great point. If you sign a bunch of these deals worth millions of dollars, then you are pulling in more money than obviously if it was an exclusive one. But according to various reports, there's some numbers on how much they've been pulling in. So the Associated Press, so the open AI working with them for their text archives for training, that's a single digit millions per year deal. The FT. Financial Times whose deal includes display and training is worth five to 10 million per year. Our parent company, Axel Springer, signed a three year deal, which includes use of its content for both training and display worth about 25 to 30 million in the biggest one I could find. News NewsCorp, Danny, you might have written about this five year deal with OpenAI valued it as much as 250 million in cash and OpenAI credits. So that's 50 million per year. They could make monies from these, Danny. But my other question would be how long will AI companies need publishers data for?

Daniel Konstantinovic (10:25):

Yeah, I mean the question is if they really needed the stamp of approval to begin with, because these AI companies are pretty notorious for having already harvested all of this stuff. I mean, we had a couple stories on the site about a database of a stolen books basically called Books three that was used by Google and Meta and OpenAI to train large language models. And the knowledge of this database sparked a bunch of lawsuits from authors and publishing companies. So the real value of these deals, I think for the AI companies is avoiding future litigation. They're already facing so many copyright lawsuits that adding to the pile would be a problem for them, and this gives them a pass on that. You're unlikely to see the Atlantic suing open AI for copy infringement now that they've signed a possible multimillion dollar deal with them. I think also, we talked about the size of the NewsCorp deal, these really large media

conglomerates that have a bunch of publications underneath them like Axel Springer, like NewsCorp or Vox, another recent one that signed a deal they stand to benefit the most from these deals because of the volume of training material that they have on offer.

Daniel Konstantinovic (11:44):

But I dunno, I agree with your sentiment, Marcus, that this does just feel like a quick opportunity for a influx of cash at a time. And in the long run, I don't think these deals will really benefit publishers.

Gadjo Sevilla (11:58):

I think they do help to legitimize AI companies who can claim that their training data comes from vetted sources. I think what's important to point out though is that quality content is a finite resource, and so at some point all the basis will be covered. And the way these training models just suck in all this information, this is not going to go on forever clearly. So I think you will see more of this in the short term. I don't know how long these deals are for or whether they're potentially recurring, but I think it's important for AI companies now as well as for publishers who, again, they don't want to miss out on any opportunity at this point to showcase their content before it gets subsumed whether they want to or not by AI engines. Yeah,

Marcus Johnson (12:53):

Yeah, because open AI's defense here is that they're saying they're simply making fair use of publicly available data. And so I'm sure that term is going to come up a lot in this case of the New York Times. What does fair use mean in this instance?

Daniel Konstantinovic (13:05):

Yeah, I mean, I think Goddard just said something really interesting, which is that this is not going to go on forever and that high quality content is a finite resource. A lot of people have posited that these AI products are set to really disrupt search, which is the way that the majority of publishers get a large portion of their traffic, and a lot of those publishers are likely to go out of business when the traffic goes away because there will be less viewers and a devaluing of ad space. And even the proposal for how AI could benefit the news business, which is largely that it could offer new ways to advertise to users or identify untapped groups of users to show ads to. Even that is not really a long-term solution because the traffic is likely

to decline. And so when all of these publications start to fold, what will OpenAI have to train its model on? The idea is that this will, it's kind of a zero sum game where eventually there will be nothing new to input into the large language models other than itself

Marcus Johnson (14:14):

Kind of a SNA eating its own tail. Yeah, I mean in terms of publishers, some publishers being hugely affected by this, there some numbers on publisher traffic and how it'll likely suffer the December internal report from the Atlantic, including large publishers could lose 20 to 40% of their traffic. This was if Google rolled out its search generative experience, similar concepts, you type something in and you get an answer back and you don't need to go look for the information by clicking on links. And so that was one data's point, the March report from RIV estimating publishers could lose as much as 60% of their traffic as a result of the changes as well. But you talked about quality content that these deals do depend on the size, quality, and type of content. Ms. Fisher of Axios noting that news publishers with large video archives think broadcasters and cable networks will have more leverage than text-based businesses because video archives are less available for bots to scrape.

Marcus Johnson (15:12):

Daniel, I think you touched on this. This is an interesting point, which is why these companies are struggling to provide a unified front, and Ms. Fisher was noting that music and book publishing both able in their fight for copyright protection to come together and actually dot dash Mary, this parent a c was pushing to create an industry coalition to help unite big publishers in fighting for copyright protections from AI companies. But that effort fell through due to conflicting business incentives supposedly. So why has New York Times gone in that direction and most other publishers gone in the other one?

Daniel Konstantinovic (15:43):

Well, the New York Times sued open ai, but not for lack of trying to strike its own deal. The lawsuit was the result reportedly of a breakdown in negotiations that the New York Times couldn't and OpenAI could not find terms that they agreed on. So the New York Times and a deal as just a lesson pointed out, is still a likely outcome of this litigation in a possible settlement. So why didn't publishers band together the way these other industries have? There's a couple reasons. I mean, there's not too much of a precedent, I suppose, for publishers backing each other in a legal issue in this way. But I'm sure that these publishers,

many of them who are quite strapped for cash at the moment probably did some simple balance sheeting of can we afford to go into a months long court battle with some of the wealthiest companies in the country and likely came out with the answer that no. And even if many of them had banded together, I'm sure that they still would not have had the deep pockets to fund that kind of litigation.

Marcus Johnson (16:52):

These do feel like kind of settlements like settlement out of court type deals. They like, we'll take the 250 million now, we might get nothing later if we lose. So let's take off the lawyer hats of this last question, gents, and tell me how you guys personally think this stuff plays out. Go to you first.

Gadjo Sevilla (17:11):

So one analogy that comes to mind was Napster and iTunes. So people were illegally sharing music and then along came a business model that promised to allow publishers, musicians to profit from the technology. Hindsight, they didn't really profit that much. \$1 songs really didn't change the game for anyone. And depending on who you ask, that pretty much killed the music industry's model from that time. Music publishing is this similar. Well, in a way, they're already taking the information. I mean, you can have say Chad, GBT, bring out results from websites from behind their paywalls. How do they do that without subscriptions? Right. So I think for publishers in this sense, they're just trying to find a balance and a way to get something out of it while still maintaining a veneer of we're progressive, we're not behind the technology cycle. Right? We're building it alongside these companies.

Marcus Johnson (18:19):

Right. One of the comments I did see was what we start to see a marketplace pop up to help companies sell their content libraries. And Tobit is apparently building a marketplace to connect AI bots and scrapers with publishers verified content for a dynamic fee. Fox Corp also building a similar tool. And so you might see marketplaces where these publishers can sell their content libraries to AI companies. Question there, Danny though, is even if there is a marketplace, how do you agree on the price of news, which is a whole other kind of worms, but Danny, how do you see this playing out?

Daniel Konstantinovic (18:53):

Yeah, I mean even news publications can't agree and consumers can't agree on what the price of news is. I wish that I had a more optimistic view of all of this, but I feel quite negatively about all of it. And I think that it's just a continuation of the same decline of the news industry that we've been seeing for the last several years. A lot of publications are really threatened by what OpenAI and others are developing and with good reason. And the largest publications like the New York Times or NewsCorp are going to be the last ones left standing because they have the largest market share, the largest audiences, the largest content libraries to negotiate with if they want to strike these kinds of deals. I think what we will see is that companies that have really made an effort, not just recently, but in the past several years or longer to develop different revenue streams or diversify their businesses are the ones that are really going to survive or coast through this period.

Daniel Konstantinovic (19:55):

So the New York Times, I'm sure we'll do well because they have a very developed subscription offering with games, with cooking with the athletic that has done really well for them. I'm sure NewsCorp will continue to do well and someone who I think has had some really smart things to say about all this is Eli Patel, who's the editor in Chief of The Verge. There was a recent episode of his podcast that talked about this concept of Google Zero and how publishers who are reliant on Google for Traffic are not unlike an influencer who is solely reliant on the TikTok algorithm for traffic. And that companies that find themselves now in that position and did not develop a of their own revenue are kind of too late to adapt to the changing industry.

Marcus Johnson (20:53):

On that point, I mean, missed lesson from The Atlantic was saying, why would anyone want to read a bunch of news articles when an AI could give them the answer? Maybe with a tiny footnote crediting the publisher that no user will ever click on. She says, tech companies aren't in the business of news and they shouldn't be. Publishers have to stop looking to them to rescue the news business. And then just final point from me, you were talking about the wealth biggest publishers being the ones that are going to get through this also seems like only the kind of rich AI companies are going to be the ones that survive, the ones who can afford to pay these publishers for their data to train their models. Who will be the ones who will survive? Because if they don't, your AI model won't be as good as the others gents. That's

why we have to leave the conversation today. But thank you so much for your time. Thank you to Danny. Yeah,

Daniel Konstantinovic (21:42):

Of course. Always happy

Marcus Johnson (21:43):

To be here. Thank you to Gaja. A pleasure. Thank you. And thank you to Victoria who edits the show, Stuart and Sophie, of course, who help out with the podcast. And thanks to everyone for listening in. We hope to see you tomorrow for the Behind the Numbers Weekly we listen, an e-Marketer video podcast made possible by round out.