

MIT CSAIL Alliances | Click Here Edited Podcast for CSAIL Alliances Export 4

Welcome to a special bonus episode from MIT'S Computer Science and Artificial Intelligence Labs Alliances podcast. I'm Kara Miller.

[MUSIC PLAYING]

This week we've got a show for you from our friends at Click Here, a podcast that tells true stories about the people making and breaking our digital world. The episode is about these new generative AI models. They know how to write, how to compose music, even how to create works of art. But they learned how to do all these things by training on data created by humans and some organizations are striking back.

The New York Times suing.

Microsoft and ChatGPT maker OpenAI over copyright infringement.

It's the first major American media organization to take action against the artificial intelligence platforms.

The show you're about to hear which debuted earlier this year, touches on something that every AI expert and every executive seeking to implement AI thinks about, legal liabilities. And as independent artists and giant media companies take their cases to court, the question is, will all this fundamentally change the human AI relationship? Here is host Dina Temple-Raston.

The cases have become a test of whether tech giants like OpenAI and Microsoft should be allowed to train its AI models on other people's content and then essentially repackage that content and sell it themselves.

Every 5 or 10 years when there's a big technology shift, there's typically a lawsuit that goes with it.

And while the *New York Times* lawsuit is the one that's getting most of the headlines, there are a roster of other suits that don't just relate to generative AI training on text, but on art and music compositions as well. The lawsuits are seeking to answer a fundamental question about generative AI, what do human creators own? And if tech companies are using their content to train AI models, shouldn't the humans be paid for it?

And our story begins not in New York, but in Nashville, Tennessee, with an artist and illustrator named Kelly McKernan.

I work in watercolor and acrylic gouache, and it's inspired by Art Nouveau, which was a movement from about 100 years ago.

Kelly, who uses the pronoun they, specializes in a magical realism. Their paintings remind me of old Grateful Dead albums like *American Beauty* or *Anthem of the Sun*. But instead of skulls and skeletons and flowers, Kelly uses the female form woven into intricate patterns and beautiful colors.

Most of my work is an extension of myself. They're self portraits, even if they're not my face.

And like most artists these days, Kelly has put their images online, not just to share it, but maybe to drum up some new business too. But then this crazy thing happened. A couple of years ago Kelly was suddenly--

Tagged in these posts on Twitter and Instagram.

--so they just clicked on one of the links and it took them to a page that had all these images that were purporting to be Kelly McKernan's but Kelly had never painted them.

Didn't look like my work, but it was using my name. So I began to look into what that was all about, why is my name in part of these images that I didn't make?

It didn't take long for Kelly to figure out what happened.

And I discovered that more than 50% of my paintings had been scraped to use as training data for AI image generators.

An AI image generator, like ChatGPT for art. In Kelly's case, a company in Sweden called LAION-5B had scraped the internet and helped themselves to all material, including Kelly's artwork.

Did that database ask you if you wanted to be part of it?

No.

But it went beyond just scraping illustrations without permission. People using LAION-5B could create AI generated images using Kelly McKernan's name.

So let's say somebody wants to create a portrait of Madonna in the style of Kelly McKernan so that is a prompt. And then these programs will generate several images created with my style based on these 50 or so paintings that were scraped. I felt violated, their bastardizations of everything I've ever made because there is no intent, there's no meaning, it's just generated slop.

It looks pretty and it looks good enough. A big part of the problem is if somebody can just generate something that looks pretty and good enough, they're not going to hire me.

Each time a stranger created a new work of art using Kelly's name as a prompt, that was one less illustration needed from the real Kelly McKernan. And Kelly wasn't going to take this lying down.

I started making a fuss about it on Twitter, talking to other artists.

And this wasn't just about losing business. These soulless AI Kelly's were affecting the real life Kelly and made them not want to--

Paint anymore because I feel so grossed out by this.

And then in December 2022, after months of watching a generative AI program offering up what looked like cheap versions of their work, Kelly wrote a post trying to get people to understand what all this was doing to them.

It basically was me being like, hey, I'm a small artist. I'm a single mom, but this is really not OK with me. This is affecting my work, affecting my career.

The post went viral and it caught the attention of another artist, a woman named Karla Ortiz, and that changed everything for Kelly.

She's well-known for her advocacy work, and she mentioned that she was a part of gathering artists for a potential class action lawsuit. She asked if I was interested, and I was like, absolutely.

Most of the discussion about generative AI these days is centered not on art, but on programs like ChatGPT or Microsoft's Copilot, which vacuum up text files from the internet to train their AI. And they've gotten away with doing that because writers have had a hard time showing they've been financially affected directly by AI programs training on their work. But that's precisely what sets Kelly's case apart. Kelly says the artists in their class action suit can document how much this is hurting their bottom line.

I typically pull in about 32 a year.

They mean 32,000.

This year I believe I made 22. I'm not making a living anymore off of my work.

Maybe claiming to have lost a third of your income to AI might feel a little crazy. But a few months ago, if you just typed Kelly McKernan artwork into a Google search engine.

And then the top result isn't even my art, it's art made from my art.

What Kelly saw at the top of that search is an image made from an AI program called Midjourney.

With artificial intelligence using Midjourney, you're able to create what you have in your mind and transfer it into a tangible work of art.

In other words, it can create images from natural language descriptions. You just prompt it with a name or a description and well, it'll give you an illustration. And in Kelly's case, an image that echoes their work. Midjourney isn't the only program that does this. OpenAI has something called DALL-E that does the same thing.

And Kelly has seen these AI Kelly-like works on stability AI's Stable Diffusion platform too. And all these AI Kelly's are part of the reason, Kelly says, that they had to get a second job.

I'm teaching right now. I'm an illustration professor at adjunct right now so I can help make ends meet because I've been losing work to the AI. And my students, they're pretty concerned. I'm concerned for them, their entry level gigs are all disappearing because it's being taken up by these programs.

Kelly ended up joining that artist who saw her viral post in a class action lawsuit against a handful of these companies, including Midjourney and Stability AI. It's essentially a copyright suit. They're hoping to either make the tech companies stop training their AI models on their art or at least pay them when they do.

Our assumption is that we are being infringed and that's the whole lawsuit. But I can confidently say I feel very infringed.

A San Francisco Federal judge assigned to the case last year voiced some doubt about whether a style can be protected by copyright. But Kelly says the fact that people use their name as a prompt suggests that there is something uniquely Kelly McKernan about what people are creating with AI and Kelly should have control over that.

My name has just been taken out of my own control. Such a bizarre thing to search my name and see AI generated art representing me that I didn't create.

Which brings us to the case that made all those headlines, the *New York Times* decision to sue OpenAI and Microsoft. And their claim is a little more, well, black and white. That's because OpenAI has admitted to training its AI model, not just on a few *New York Times* articles, but on its entire archive, millions and millions of pieces dating back to 1851.

The Times contends that Microsoft and OpenAI seek to free ride on *The Times's* massive investment in its journalism. *The New York Times* isn't seeking any specific damages, but it does want OpenAI and Microsoft to destroy any ChatBot models that incorporate their material.

[INAUDIBLE] is a lawyer who specializes in tech-related law, and she says *The Times* case has an advantage over cases like Kelly's.

Some of the art cases, it's not as clear that the output is so obviously a possible substitute for the original.

But in its complaint, *The New York Times* says it has proof that that's exactly what's happening.

They have this amazing exhibit, Exhibit J. There's 100 instances of requesting something from ChatGPT and ChatGPT spits out effectively verbatim *New York Times* articles on some really interesting topics.

[INAUDIBLE] has been working in tech for a while. She was one of the founding members of Amazon's Alexa team, and now she's the CEO of an AI startup for lawyers called General Counsel AI. She says that Exhibit J is the smoking gun content creators have been hoping for.

And the *Times* case goes beyond just appropriation of reporter's work. They claim AI is causing reputational damage as well by attributing things to the paper that the AI model is just making up.

If you ask ChatGPT what the *New York Times* says about orange juice causing lymphoma, it will actually give you a response and say--

A wrong response known in AI circles as a hallucination.

--ChatGPT confidently responds. *New York Times* published an article on Jan 10, 2020, titled Study Finds Link Between Orange Juice and Lymphoma. Of course, this is all completely false, it's generated from whole cloth.

AI's hallucinations aren't just an unfortunate glitch, *The Times* argues. It's harming the gray lady's brand. Microsoft and OpenAI didn't respond to our request for comment. The case is expected to go to trial in 2026, and Julia says, it isn't clear how it'll turn out.

It's not a slam dunk on either side. I would say, possible outcomes include saying, OK, it's fair use if you have certain protections against the verbatim spit backs.

Or they might work out a licensing agreement or Congress could step in. In fact, Congress already has.

This is our second hearing in as many months on the intersection of artificial intelligence and intellectual property law and policy.

But Julia says all this focus on copyright isn't going to spell the end of AI.

This technology is too useful and the genie is too much out of the bottle for it to be enjoined or in any way stopped.

AI's got something to say.

Looking back on it now, it's almost inevitable that these generative AI models would raise copyright concerns. And though AI is the new, new thing, copyright law is quite literally as old as America itself.

Copyright literally comes from the constitution to promote the progress of science and the useful arts. Congress may secure limited time monopolies for authors and creators.

That doesn't mean you can't use published works. You can if you abide by a doctrine known as fair use. And to avoid running afoul of copyright, at a minimum you need two things. First, you have to transform the copyrighted work in some fundamental way. And second, and this part is key, whatever you're creating can't be used as a substitute for the original work.

So if, for example, a critic quotes passages from a book in her review, it's probably considered fair use because it creates a new work. But if a generative AI model like ChatGPT is training on a bunch of articles and then just spits out something verbatim, well, that could be a copyright problem.

When generative AI first came on the scene, did you think that there was going to be a problem with copyright?

I would say yes, only because vast amount of data that in this case GPT is trained on, I think it's equivalent to 65 libraries of Congress.

65 libraries of Congress, think about how much that is and how could you ask permission for all of that? That may be why AI companies decided to ask for forgiveness instead. The sheer volume of data they needed to train their AI would make getting permission pretty hard.

So instead they're helping themselves to troves of information under the belief that they are, in the words of the constitution, using it to promote the progress of science and useful arts. The tech companies openly admit that their training models on the *New York Times* articles, but their argument is--

But we're not using it to read it, we're not using it as a substitute for a subscription, instead, we're using it to create this entirely new technology. The magic word is transformative.

Where you draw the transformative line is tricky, even for people who saw the copyright infringement problems coming, like this guy, Ed Newton-Rex.

I've been in generative AI for 13 years before generative AI was cool I guess.

Until recently, Ed led the audio team at an open source generative AI company called Stability AI. Yes, the same company that was scraping Kelly McKernan's illustrations from the internet. But Ed wasn't there as your average tech pro.

I'm, I would say, a better musician than I am a computer guy. I've done a bit of coding.

Back in 2012, Ed founded an AI music generation startup called JukeDeck. And even then he was wrestling with these copyright issues as they relate to music. And more broadly, the iterative quality of creativity.

I want to start with a quick quiz, who can tell me which song these chords are from. [MUSIC PLAYING]

He talked about doing this during a Ted Talk in London.

Any answers? Journey, Don't Stop Believing. Exactly that is one correct answer. But there are actually a few other answers you could have given. This is one, [JAY-Z AND MR. HUDSON, "FOREVER YOUNG"] Forever young, I want to be forever young.

What he's pointing out here is that human creativity doesn't just appear out of thin air, it's iterative in a way that's not dissimilar from AI. The same chords can be used to make many songs. The main difference, of course, is that we might listen to tens of thousands of songs or read several thousand books in a lifetime, while AI is gobbling up 65 libraries of Congress. That's over 3 billion books in just one fell swoop.

AI's iterative capacity is infinite, and ours not so much.

(SINGING) Napster to go, go, go, go, go, go.

The AI copyright issues we're wrestling with today Ted told me remind him a little of what happened with Napster.

Napster introduces Napster To Go. Fill your MP3 player with an unlimited number of songs to try it for free.

Napster, you remember, was the website from the early 2000 that allowed people to download loads of music for free. It had millions of users until bands like Metallica and artists like Dr. Dre sued it for copyright infringement. Napster went bankrupt in 2002.

In that case, though, they weren't iterating on people's music. They were just stealing it. But the reason Napster has a role in this AI copyright problem is because of what rose from its ashes, something that allowed content creators to get wide distribution and get paid. And that, of course, was iTunes.

This is from an early iTunes App. Ed Newton-Rex thinks generative AI is headed in the same direction to this licensing model.

[JET, "ARE YOU GONNA BE MY GIRL?"] One, two, three tap my hand and come with me because you look so fine that I really want to make you mine.

I think we will graduate out of this Napster era, the Wild West era where people are doing what they want in regards to training data. And I think we will gradually move to more of the iTunes model where licensing prevails because I think that--

Ed already knows what a world where licensing prevails would look like, because when he was building his music making AI back in 2012, he did something that now seems quaint. He actually asked artists for permission.

We licensed all of the training data, all of the content, all of the music that went into that. And so that was exciting because that was the first really large music model that had licensed training data in that way.

Of course, getting permission to do all that took time and money. But Ed, you might recall, is a musician, so it was important to him that credit be given to the content creators that were helping him build Jukedeck.

But it's something that I've always been incredibly passionate about really as a composer I think it's really, really important for AI systems, AI models, and AI companies to work in a way that really respects creators rights. And I think the way you do that is by partnering with them to get their training data and not simply taking it.

Remember, after Ed sold Jukedeck, he went to work for a generative company called Stability AI, where he was in charge of its audio team. And last fall he left the company precisely because he didn't agree with the practice of just taking data without paying for it.

Yeah, that's why I stepped down from Stability, not so much in protest at Stability itself and their practices, but more them as part of the wider industry. I think lots of generative AI companies, big generative companies take the same approach. The Stability too do, which is generally an approach that says, well, it's fair use of copyrighted material to use it without consent. And I don't agree with that. So that's why I stepped down.

We reached out to Stability AI for comment on both Ed's departure and the artist's class action and got no response. Earlier this month, Ed decided instead to start a nonprofit that would help consumers see right away whether the AI they're using is paying creators their fair share.

So we're essentially going and certifying AI models who take a more ethical approach to how they train and how they reward and how they get consent for and from their creators.

It's called fairly trained, and it's like a fair trade certification, except instead of coffee, it's AI models.

And I'm hoping that by doing that we give this label to a companies and people and companies out there who do care about where training data's come from can go and look at this list and say, well, I'd rather use one of these models.

You can identify AI that aligns with your values and then support it. But that may be a long way off. So in the meantime, creators like Kelly McKernan, the artist we met at the beginning of the episode, are finding ways to fight fire with fire right now.

They're starting to incorporate something called data poisoning into their work. They're tweaking the pixels in an image in a way that humans can't see but confuses a generative AI model. There's one program that does this called Glaze and another called Nightshade.

I actually have a piece in their lab right now that I just finished that they are adding Nightshade to. I'm very excited to be one of the first artists to get to use it.

They essentially trick the AI into seeing something that isn't really there, a electronic mirage, a watermark on steroids. In other words, using tech to fight tech until the humans figure out how to get this right.

This is [Click Here](#). And this is MIT'S Computer Science and Artificial Intelligence Labs Alliances podcast. I'm Kara Miller. Thank you so much for listening to our bonus episode from the podcast [Click Here](#).

We will be back soon with another brand new episode of our own. Until then, you can find out more about CSAIL latest research and the CSAIL alliances program at our website, cap.csail.mit.edu. I'll talk to you soon.

[MUSIC PLAYING]