

AI, Originality, and Creativity: Copyrighting All the Melodies to Avoid Accidental Infringement

*Damien Riehl**

If you think of the “Live Your Life” song, who thinks that “Levitating” by Dua Lipa sounds like—or is substantially similar to—“Live Your Life,” by Artikal Sound System? Who does not think they sound substantially similar?

It’s a trap. That’s the wrong question to even ask. Really, the question of substantial similarity is number three in this. Really, the order of questions we should be asking is the following: is this melody even copyrightable in the first place? That’s the first question. And that’s what we are going to be talking about today. In 2019, I said to my colleague, Noah Rubin, “Should we break music?” And he said, “Heck ya, let’s break music.” And that’s how the *All the Music Project* came about.

I’m going to tell you a true story, but instead of the name of the protagonist, think about your favorite artist. Think about your favorite musician and think about your favorite song by that musician. Think about them bringing that song from nothing to something into your ears and bringing you so much joy.

Now think about your favorite musician getting sued and that lawyer saying to your favorite musician, “I represent this group. I think you heard their song and then you wrote yours. You infringed their copyright.”

And imagine your favorite musician saying, “No, it’s not true. I don’t think I’ve ever heard that song. But even if I did, I certainly wasn’t thinking about them when I made my song.”

Imagine the case going to trial and a judge saying, “I believe you; I don’t think you consciously copied that group. But what I think did happen is you *subconsciously* copied them. You infringed their copyright, and you have to pay them a lot of money.”

* Technology Lawyer, VP at vLex, SALI leader, and Founder of *All the Music, LLC*. This is an edited version of a keynote speech given at a symposium entitled “Rhythm, Rhyme, and the Rule of Law” at Chapman University Fowler School of Law on February 2, 2024.

Think about whether that's fair or just. This actually happened to George Harrison, the lead guitarist of The Beatles, and the group was The Chiffons, who had a song that goes, "He's so fine, oh so fine." And George Harrison had a song that goes, "My sweet Lord, oh, sweet Lord." [Sings the same melody.] But what neither George Harrison nor The Chiffons nor the judge, nor anybody else had really considered, is, maybe, since the beginning of time, the number of melodies is remarkably finite. Maybe there are only so many melodies in this world. And The Chiffons, when they picked their melody, plucked it from that already existing finite melodic dataset. And George Harrison happened to have plucked the same melody from that same finite melodic dataset.

When many laypeople think about musicians, they think about them drawing from their own creative wellspring, bringing from nothing, something into the world. They have a blank page upon which they can put their creativity. That's actually not true. As George Harrison realized, you have to avoid every song that's ever been written because, if you don't, you get sued. If you're lucky, you pluck one of those already existing melodies that hasn't been taken. If you're unlucky, you pluck a melody that's already been taken—whether you've heard that song or not. Maybe you've never heard it before. If that happens, if you're lucky, you have a co-songwriter or somebody else who says, "That new song sounds a lot like that old song." And you change it before it goes out the door. Now, if you're unlucky, you don't have somebody telling you that; you release it out in the world, the group hears your song, and they sue you for a song that you've maybe never heard before in your life. You've just stepped on a *melodic landmine*.

The thing is, this is the world before my colleague, Noah Rubin, and I started our project. The world now looks like this: we filled in every melody that's ever existed and ever can exist. Every step is going to be a melodic landmine. And ironically, this is actually supposed to help songwriters. Let me tell you how.

I am a lawyer and have been since 2002. I have litigated copyright cases and also taught copyright in law school. I am also a musician. I have a bachelor's degree in music. I'm a performer and recording artist. And I also produce records. I am also a technologist. I've been coding since 1985, for the web since 1995. I've done cybersecurity and I also currently design software. So, that puts me right in the middle of a Venn diagram that gives me a few insights that, if I were in any one of those areas, I might not have had. And my colleague Noah Rubin, in addition to being one

of the smartest people I've ever known, is also a musician, and he's also one of the most brilliant programmers I've ever known.

Between our work, we came to a realization that you may have been had: "You know that new song? It sounds a lot like this other old song." And there's a reason for that. We've demonstrated that there are only so many melodies. There are only so many notes that can be arranged in so many ways.

And that's different than visual art, where there are an infinite number of brushstrokes, colors, and subjects that, to accidentally mimic them, is very difficult. Similarly, with language, the English language has 117,000 words in it, so the odds of accidentally writing the same paragraph are next to zero.

In contrast, music doesn't have 117,000 words. Music has eight notes: *Do, Re, Mi, Fa, Sol, La, Ti, Do*. One, two, three, four, five, six, seven, eight. And every popular melody that has ever existed and ever can exist is those eight notes. Now, it's remarkably small.

I worked in cybersecurity, and I know if I wanted to attack your password and hack your password, one way to do it is to use a computer to write really quickly "AAA." No? "AAB," "AAC," and to keep running until it hits your password. That's called brute-forcing a password. So, I thought, what if you could brute force melodies? What if you could say, *Do-Do-Do-Do, Do-Do-Do-Re, Do-Do-Do-Mi*. And then exhaust every melody that's ever been. And the way the computer reads music is called the Musical Instrument Digital Interface ("MIDI"). And in MIDI, it looks like this: *Do-Do-Do-Do, Do-Do-Do-Re*.

So, I asked my colleague Noah, asking "Can you write an application to be able to march through every melody that's ever existed and ever can exist?" He responded, "yeah, I could do that."

So, at a rate of 300,000 melodies per second, he wrote a program to write to disk every melody that has ever existed and ever can exist. And the thing is, to be copyrighted, you don't have to do anything formal. As soon as it's written to a fixed, tangible medium, this hard drive is copyrighted automatically.

Now, this leaves copyright law with a very interesting question, because you think about the world before and songwriters had to avoid every song that has ever been written, in

red.¹ Noah and I have exhausted the entire melodic copyright. So, if you superimpose the songs that have been written, in red,² with the songs that haven't yet been written, you have an interesting question: have we infringed every melody that has ever been? And, in the future, every songwriter that writes in the green spots,³ have they infringed us?

Now, you might think at this point: are you some sort of copyright troll that's trying to take over the world? And I would say, "No, absolutely not." In fact, the opposite is true. Noah and I are songwriters ourselves. We want to make the world better for songwriters. So, what we've done is we've taken everything and put it in the public domain. We're trying to keep space open for songwriters to be able to make music. And we're not focused on the lyrics. We're not focused on recording. We're focused on melodies. And the thing is, we're running out of melodies that we can use. The copyright system is broken, and it needs updating.

Some of the insights that we've received as part of our work are that melodies, to a computer, are just numbers because those melodies have existed since the beginning of time, and we're only just discovering them. So, the melody, *Do-Re-Mi-Re-Do*, to a computer is literally 1-2-3-2-1. So, really, the number 1-2-3-2-1 is just a number. It's just math that has existed since the beginning of time. And under the copyright laws, numbers are facts. And under copyright law, facts either have thin copyright, almost no copyright, or no copyright at all. So, maybe if these numbers have existed since the beginning of time where we're just plucking them out, maybe melodies are just math, which is just facts, which maybe are not copyrightable. Maybe if somebody's suing over a melody alone—not lyrics, not recordings, but just the melody alone—maybe those cases go away. Maybe they get dismissed.

Now you might ask, "Well, what constitutes a melody?" And we were initially going to take the entire piano keyboard and do the entire keyboard. But we thought, let's focus on the vocal range, which is actually two octaves. And then we thought, no, actually we're talking about pop music, which is the only thing that makes money that people sue over. So, we looked at musicologists, and they have debated what is a motif (a short melody) versus a longer

¹ For a visual representation, see TedxMinneapolis, *Why All Melodies Should Be Free For Musicians to Use*, TED (Aug. 2019), https://www.ted.com/talks/damien_riehl_why_all_melodies_should_be_free_for_musicians_to_use?language=en [<https://perma.cc/SR6A-4D8H>].

² *Id.*

³ *Id.*

melody. And we landed with twelve notes. And then we superimposed that number with songs that have either been litigated or threatened to be litigated.

The Chiffons' "He's So Fine" goes [singing], "He's so fine, oh so fine." That's eight notes. And then George Harrison's "My Sweet Lord" goes [singing], "My sweet Lord, oh sweet Lord."⁴ Also, eight notes. So, okay, cool, let's go with "Under Pressure" by Queen versus "Ice Ice Baby" by Vanilla Ice.⁵ Super close, right? Tom Petty's "I Won't Back Down" goes [singing], "Oh, I won't back down, no I won't back down." Ten notes. Sam Smith's "Stay With Me" goes [singing], "Won't you stay with me 'cause you're all I need."⁶ Ten notes. Think of Flame's "Joyful Noise" versus Katy Perry's "Dark Horse." Different melodic ending? Jury didn't care. Verdict for \$2.8 million.⁷ That melody shows up in our first dataset 8,128 times. So, Flame—who sued Katy Perry—should he get a monopoly for life of the author plus seventy years over that thing that showed up 8,000 times in my dataset? My machine cranked it out at 300,000 melodies per second. Should he get a monopoly for life of the author plus seventy years? That's the real question. That's the crux of this issue.

So, dataset number one: major scale and minor scale. Dataset number two: many said, "*Um, actually* you are not covering jazz and you are not covering classical." Cool. We'll do chromatic too now. Now we've got the western scale. "*Um, actually* you're not really covering rhythm." Cool. We'll do rhythm too. We'll add one note that's called "silence." Now we've got every rhythmic melody. Any variation on this is just math. Give me another, "*Um, actually,*" and I'll give you another dataset. This is all math.

So, Noah, who worked with me in cybersecurity—best cybersecurity coder I have ever seen—he and I are going to be speaking at the South by Southwest conference ("SXSW") in

⁴ See *Bright Tunes Music v. Harrisongs Music*, 420 F. Supp. 177 (S.D.N.Y. 1976).

⁵ See Stephanie Sengwe, *Musicians Who Were Accused of Copyright Infringement by Other Musicians: Can You Hear the Similarities?*, PEOPLE (May 5, 2023, 12:20 PM), <https://people.com/music/musicians-accused-of-copyright-infringement-by-other-musicians/> [<https://perma.cc/JNX4-44UM>].

⁶ *Id.*

⁷ The judge in that case later overturned that verdict, explaining that the melody was not "particularly unique or rare." A three-judge panel for the Ninth Circuit later affirmed the district court judge's decision. See *Katy Perry's "Dark Horse" Comes Out in Front: Ninth Circuit Affirms No Infringement in Copyright Lawsuit*, JD SUPRA (Mar. 16, 2022), <https://www.jdsupra.com/legalnews/katy-perry-s-dark-horse-comes-out-in-8093022/> [<https://perma.cc/EEA4-5JKG>]; Mark Savage, *Katy Perry Wins in Dark Horse Copyright Appeal*, BBC (Mar. 11, 2022), <https://www.bbc.com/news/entertainment-arts-60705977> [<https://perma.cc/B2FQ-NZ23>].

Austin, Texas.⁸ There, we are going to be speaking together and jamming. He's a brilliant guy who now works for Amazon Web Services. If you were to use Amazon, you have him to thank for keeping you safe. He created this dataset for All the Music. We put everything in the public domain. It is all on GitHub. You are able to do it right here. If you want to join us and if you want to help us, go ahead, and do that. You can expand our dataset beyond what we have done to the entire keyboard with every rhythmic variation, and it's all math. We're just going to exhaust all the things. So, want to help? Go ahead and join us.

Really, the other question, though, is we have talked about the red copyrighted spots:⁹ What about Bach? What about Mozart? So, if the red spots¹⁰ are copyrighted, what about the gray spots¹¹ that are already in the public domain? There's a real question: can I take a Bach melody and pull it out of the public domain? Or is what's in the public domain always in the public domain, whether you've heard Bach or not? What if you were to superimpose all the copyrighted spots on the public domain spots? And that's what I actually am going to be doing.

Spotify has a patent, saying, "If you input a melody, we will tell you every song where this melody shows up." And so I have a buddy who's the former Chief Economist of Spotify, and I said, "Hey, can you connect me with that inventor?" And he said, "Sure." So I'd like to take all 471 billion of my melodies and put it into Spotify's algorithm, to be able to say: where are the open spots and where are the closed spots? And more importantly, where is Bach? Where is Beethoven? Where are these things that maybe somebody thought that they had copyrighted the melody, but it has been in the public domain since the 1700s?

The thing is, were running out of melodies. If you think about a song that has an intro, verse, pre-chorus . . . there may be a bunch of—maybe ten—melodies in a particular song, countermelodies, etc. Streaming services currently have an estimated 375 million songs, and they'll likely have 450 million songs by the end of the next year. At ten melodies per song, that's four and a half billion melodies that are just being cranked out.

⁸ South by Southwest celebrates the convergency of tech, film, music, education, and culture. See SXSU, <https://www.sxsw.com/about/> [<https://perma.cc/CW96-XQUD>] (last visited Apr. 14, 2024).

⁹ See TedxMinneapolis, *supra* note 1.

¹⁰ *Id.*

¹¹ *Id.*

There are only so many notes: *Do-Re-Mi-Fa-So-La-Ti-Do*. That's sixty-eight billion melodies, that's eight to the twelfth power. So, sixty-eight billion melodies, and with our most-recent dataset we've already hit over 470 billion melodies. And let me tell you, of those sixty-eight billion, a lot of them sound awful, like, *Duh, duh, duh, duh, duh, duh, duh*. Nobody's going to listen to a melody like that.

The number of listenable melodies is remarkably finite, and we're running out of them. Because every bedroom producer is recording songs things and uploading them to YouTube, Spotify, etc. Everyone in their bedroom is doing these things. So, the number of spots we're trying to keep open are for people to be able to make more music.

Who knows about Judge Learned Hand? Great guy. He came up with the idea of independent creation. He said that "Ode on a Grecian Urn" was a poem and he said that, hypothetically, as a thought experiment, if you're writer who came up with the exact same words from "Ode on a Grecian Urn" without having read the original Keats poem, both you and Keats would still have independent copyright in that poem—because you independently created that, since you haven't heard Keats before. That was the beautiful thing about independent creation.

But for music, independent creation is pretty much dead, or, at least, it has been dead since the George Harrison case. Independent creation really goes to "access." "Did I have access to 'Ode on a Grecian Urn' or not?" If the answer is, "Yes, I had access," then I violated the copyright. If the answer is, "No, I didn't hear it," I can independently create it and both of us can have copyright in this thing.

But the thing is that "Ode on a Grecian Urn" type of access didn't matter to George Harrison, because he said, "I didn't have access to The Chiffons—I didn't hear their song." But the court said, "No, I think you *subconsciously* infringed; therefore, whether you had access or not, you don't remember."

So, George Harrison, had to prove a negative—prove that he had *never* heard a song before. Now, where in the law have you ever had to prove a negative (which, by the way, is philosophically impossible)? You cannot prove a negative. In almost every case, the plaintiff has to prove that the defendant *did* something. The onus is usually on the plaintiff, on The Chiffons, to prove that George Harrison did the thing.

Here, the judge flipped it, saying, “George Harrison, it’s your burden to say that you’ve *never* heard it.” And how are you going to be able to say that you’ve never heard something in the grocery store? You’ve never heard something with your friend holding a phone up to you? It’s impossible to say you’ve never heard this song before.

So, this has been an injustice since when I attended Mitchell Hamline College of Law. This seems stupid. Because almost never has the onus been on the defendant to prove a negative. So, “subconscious infringement” killed Justice Learned Hand’s independent creation. Every case since that has said that you have to prove that negative, which is, again, impossible: grocery store, friend’s car. Learned Hand is sad.

But really, this goes to element two: Access. Do you have access to the thing?

When you think about access, there are cases where they’ve *definitely* had access. John Fogerty definitely had access to John Fogerty—and Creedence Clearwater Revival’s “Run Through the Jungle.”¹² This is an actual lawsuit where Fogerty was sued by his label Fantasy Records for his use of “Run Through the Jungle” in his single “The Old Man Down the Road.” Fogerty definitely had access to Fogerty.

No access is if a baby sings into a recording—and by the way, the toy’s fixed tangible medium enables copyright— [singing] “Da-da-da-da-daaa, da-da-da-da-daaa” to the same tune as Taylor Swift’s “Shake It Off,” which goes “I stay out too late. Got nothing in my brain.” No access because Taylor Swift has never heard that baby. Therefore, no access: that’s a clear independent creation, right?

But almost none of the cases are “access” or “no access”; it’s all in the middle. They’re all either: maybe you heard it, maybe you didn’t; maybe you infringed, maybe you subconsciously infringed. Almost all of these are a fact question. And the tricky part about fact questions and litigation is that most of the cases—The Chiffons’ case and Katy Perry’s case—are fact questions. And when you think about the timeline of cases, the fact question is never done at the cease-and-desist letter stage: “pay me money, or I’ll sue you.”

You go through the litigation lifespan. The fact-question lawsuit takes years. And fact questions don’t get decided until the

¹² Fogerty v. Fantasy, Inc., 510 U.S. 517 (1994).

end. And how much does it cost to get to the end? In legal fees—that’s lawyer fees—the cost is about \$2 million on the high end.

So, am I as a songwriter going to roll the dice spending \$2 million on lawyers—and then maybe pay damages on top of that—for this fact question as to whether I heard the song or didn’t hear the song? Or subconsciously infringed?

Tom Petty apparently threatened to sue Sam Smith over, ‘Oh, I won’t back down’ versus ‘Won’t you stay with me.’ Sam Smith might have thought “Wait, am I going to spend \$2 million to roll the dice? And then to be found that I subconsciously infringed Tom Petty?” Even though Smith said they’d never heard Tom Petty’s song before: Co-songwriter. Radiohead got sued by The Hollies: co-songwriter. Same song (Creep) —Radiohead apparently threatened to sue Lana Del Rey. They allegedly settled. Because nobody wants to roll the dice for a subconscious infringement to have to prove a negative that can’t be proven, right?

So, this is what really stuck in my craw in 2000 when I was in law school. I thought, this is stupid. Shouldn’t we be able to dismiss these things at the front end of the case? Saying these melodies are unoriginal, therefore uncopyrightable, and maybe the case goes away on a motion to dismiss?

Or the plaintiff should have the onus like they always have had the onus in other cases: Plaintiffs have the burden of proof. Plaintiffs should be required to show that the defendant actually had access—give evidence of *actually* having access—and absent that kind of evidence, the case goes away. That’s where the onus should be.

So, access? That’s the wrong question. Substantial similarity? Wrong question. Real question: is the melody copyrightable in the first place? Because asking if it is copyrightable first is the right thing to do.

So, if any of you are litigating these types of cases, or if any of you are a judge, look to copyrightability first. Do that analysis first. If it’s not copyrightable, do not pass go. Do not collect \$200. If it’s not copyrightable, it goes away.

Before my TED Talk,¹³ litigants and courts debated whether the alleged infringer had access and whether the melodies were substantially similar. After my TED Talk, everybody moved to question whether the melody is even copyrightable.

¹³ See TedxMinneapolis, *supra* note 1.

The “Blurred Lines” case¹⁴ essentially held that you can copyright a vibe. In the Katy Perry case, the court denied judgment as a matter of law, and then a jury verdict reached a verdict of \$2.8 million. Every defendant before my talk lost. Because everyone focused on “access” and “similarity.” Most didn’t even address originality and copyrightability. It wasn’t even argued.

Then my TED Talk happened, where I said, “Hey, these things are uncopyrightable.” My talk was released on YouTube in January 2020, and it blew up. It became viral. Within three days after it was released, it had 200,000 views. Currently, in its various incarnations on the web, it has about 2.1 million views. In January 2020, I also was interviewed by Adam Neely, the YouTuber. That interview had almost 800,000 views within a few days. We were also trending on Reddit. *The Atlantic* did an article about this.¹⁵ We were the top post on Reddit.¹⁶ We were on *Motherboard* by VICE, *The Independent* out of Ireland, and *The Telegraph* out of London.¹⁷ I was on CBS Sunday Morning.¹⁸ I got invited to a group, the “Pho List,” which is a bunch of music industry lawyers, tech people, and music tech people, where we did a lot of debates about my project.

My talk got a lot of press, and it really shifted this conversation to say, “Hey, is the melody even copyrightable in the first place or not?”

And so, after that, the Ninth Circuit, in the Led Zeppelin case, said “No, that melody for ‘Stairway to Heaven’ is uncopyrightable. Unoriginal, therefore uncopyrightable.”¹⁹ The verdict against Katy Perry was also reversed. Back in 2018, the judge essentially said, “No, I’m sorry. I can’t rule in favor of Katy Perry as a matter of

¹⁴ See Sengwe, *supra* note 5.

¹⁵ See Alexis C. Madrigal, *The Hard Drive With 68 Billion Melodies*, *The Atlantic* (Feb. 26, 2020), <https://www.theatlantic.com/technology/archive/2020/02/whats-the-point-of-writing-every-possible-melody/607120/> [<https://perma.cc/HR22-ZGXD>].

¹⁶ See u/Gustheanimal, REDDIT, https://www.reddit.com/r/MadeMeSmile/comments/15y7oig/damien_riehl_copyrighted_every_music_melody_known/ [<https://perma.cc/6FNK-H8EZ>] (last visited Apr. 15, 2024).

¹⁷ See *Computer Coders Create Every Melody Possible In Bid to Help Musicians Avoid Copyright Suits*, *THE TELEGRAPH* (Feb. 26, 2020), <https://www.telegraph.co.uk/news/2020/02/26/computer-coders-create-every-melody-possible-bid-help-musicians/> [<https://perma.cc/3AZM-MV5E>].

¹⁸ See *The Blurred Lines Surrounding Music Copyrights*, *CBS News* (Apr. 4, 2022, 4:28 PM), <https://www.cbsnews.com/news/music-copyrights-songwriting-lawsuits/> [<https://perma.cc/5589-FSTZ>].

¹⁹ *Ninth Circuit Shows Led Zeppelin a Whole Lotta Love in ‘Stairway’ Copyright Win*, *JD SUPRA* (Apr. 3, 2020), <https://www.jdsupra.com/legalnews/ninth-circuit-shows-led-zeppelin-a-10238/> [<https://perma.cc/BK78-HNF6>].

law.”²⁰ Then after my TED Talk, the judge essentially said, “No, as a matter of law, I’m going to reverse the jury verdict,” stating that the melodies are unoriginal, therefore uncopyrightable, and reversing her earlier ruling.²¹

The Ninth Circuit, then affirming, said, “as a pitch sequence, you are not entitled to copyright protection because these are the building blocks that are in the public domain.”²² That’s what I argued in my TED Talk. These are building blocks of music.

And then my friend Jennifer Jenkins asked me to speak at Duke Law. I was speaking there and, that morning, Ed Sheeran won his case in the United Kingdom (UK).²³ Back in my talk, I said, “There are only so many melodies, and we’re running out.” After he won his UK case, Ed Sheeran put out a press release, saying, “There’s only so many notes and very few chords used in pop music.”²⁴ And I said, “Dude, you not only stole the melody, but you also stole my line.”

Anyway, correlation is not causation. That is, none of them cited me, but really, the question is, do you need to cite me? Because it’s not the thing that I did. It’s the *idea* of the thing that I did that really matters. “Copyrightable?” is the question. And the real question: is Judge Learned Hand’s idea of independent creation still dead? Or does it even matter anymore? Because if you only go to the question of copyrightability, you don’t need to pass go. You don’t need to collect \$200. You don’t need to figure out substantial similarity. All that matters is, “is it original”? And if not, the case goes away.

Now we go to the machine composers. In the 1960s, specifically 1965, the U.S. Copyright Office essentially said, “You know, there’s going to be a time where the computers are going to

²⁰ See *Katy Perry Loses Summary Judgment In “Dark Horse” Copyright Infringement Case, As Court Decides That Commercial Exploitation Not Necessary to Show “Widespread Dissemination” For Purposes of Proving “Access”*, MCPHERSON LLP (Sep. 5, 2018), <https://mcperson-llp.com/katy-perry-loses-summary-judgment-in-dark-horse-copyright-infringement-case-as-court-decides-that-commercial-exploitation-not-necessary-to-show-widespread-dissemination/> [<https://perma.cc/SFL8-XJXE>].

²¹ See *Katy Perry’s “Dark Horse” Comes Out in Front: Ninth Circuit Affirms No Infringement in Copyright Lawsuit*, JD SUPRA (Mar. 16, 2022), <https://www.jdsupra.com/legalnews/katy-perry-s-dark-horse-comes-out-in-8093022/> [<https://perma.cc/EEA4-5JKG>].

²² *Id.*

²³ *Ed Sheeran Wins Shape of You Copyright Case and Hits Out at ‘Baseless’ Claims*, BBC (Apr. 6, 2022), <https://www.bbc.com/news/entertainment-arts-61006984> [<https://perma.cc/G2S9-BGRK>].

²⁴ Bill Donahue, *Ed Sheeran Heads to Trial Over Claims He Copied Marvin Gaye’s ‘Let’s Get It On’*, BILLBOARD PRO (Apr. 21, 2023), <https://www.billboard.com/pro/ed-sheeran-copyright-lawsuit-trial-lets-get-it-on/> [<https://perma.cc/26LH-J9YF>].

be able to make words. What are we going to do then? The number of words written by computers will increase, and you cannot categorically say that you *cannot* copyright those things. A computer could be used like a typewriter. And so, the question is—the human authorship—is that merely the machine being the instrument that the human is operating?” They were thinking about this in 1965. Pretty good, right?

We’re there now, right? The question is when Noah and I used a machine to brute force “*Do-Re-Mi-Fa-So-La-Ti-Do*,” and did eight up and twelve across, was that machine doing our bidding? Were we setting the metes and bounds of the copyright—and is that a “creative” aspect? Or was that just math that was unoriginal, therefore uncopyrightable? That’s the real question.

When you think of a flow chart: are machine-created works copyrightable or not? Applied to our project, if the answer is “Yes, machine-created works are copyrightable,” then we just copyrighted 471 billion melodies, putting them in the public domain. Have at it. So that’s option number one.

If you say “No, machine-created works are uncopyrightable,” the question is why? Why are they uncopyrightable? Is it because they are facts and ideas? Maybe. Is it because they’re unoriginal and uncreative? Maybe. If so, all of my 471 billion melodies are *still* public domain—because they are facts that are unoriginal and uncopyrightable.

Then, the next step is: what if a human makes that *identical* melody? Right? Like Flame’s melody, [singing] *Dun, dun, dun, dun, dun, dun, dun, dun, dun*. In my dataset Flame’s melody shows up about 8,125 times. Should Flame get a monopoly on that unoriginal melody? Does the melody somehow flip, switch—from machine-created unoriginal, uncopyrightable to somehow copyrightable—just because he plucked that same melody out of the finite dataset? His all of a sudden becomes “original”?

No, it’s stupid. All of these things—471 billion melodies—are ostensibly uncopyrightable. These cases should go away because there are only so many of those notes. Most of the people have said, even law professors say, “*Um actually*, Damien, people didn’t have access to your 471 billion melodies.” Focus on originality.

More interesting is this question: are these melodies unoriginal? TED Talks are limited to only eighteen minutes. I wanted to go deeper into originality, but the organizers said “Eighteen minutes, dude.” And I’m like, “Alright, eighteen

minutes, I'll limit originality." But this is the thing that now I'm able to talk about with you. Should we give a government copyright monopoly, life of the author plus seventy years, over something our computer spit out at 300,000 melodies per second?

With large language models, this is accelerated. I was asked by the Bench and Bar of Minnesota, "Can you write something about ChatGPT?" I said, "How long do you want it to be?" "Seventeen pages." I said, "No, not going to do it." I'm flying all over the world, talking about AI. I'm building AI systems. My rule of thumb is one page per hour, so that would take seventeen hours I just don't have. But then I realized, this is on ChatGPT, right? So like I do for every writing project, I created an outline. So, this is about three pages of really good multi-level bullets, covering, "What are large language models?" "How do they work?" "Why do they matter to the law?" "How did they beat the bar exam?"

I took that outline, and then I prompted ChatGPT to say, "Here's an outline for an article for a legal magazine. Expand it into an article. For each bullet point, give me one or two sentences." And ChatGPT spits out nineteen pages of really good stuff. I'm not done though. I spent the next three hours adding, editing, revising, putting it into my voice. I was jamming with this thing. This isn't a robot author; it's a co-author. I was jamming with it. Then, I sent the paper to the editor, who responded "Cool, let's get it out the door. I don't even need to edit it much." It took a seventeen-hour process and shrunk it down to three hours. That's a 5x increase by my math.

But the real question is: who wrote my article? That was my three pages of outlines. Those were my ideas. Could ChatGPT create those good ideas? No, because I've been thinking since November of 2022—when ChatGPT came out—every waking moment, all I've been thinking about is how large language models can affect the law. So, it was my every waking moment that created those ideas. I would argue that none of you could do that like I did that.

So, then the real point is—I asked it to do *one* expression of my ideas. I could have just as well said, "Give me a thousand expressions." "Give me ten thousand expressions." "Give me a hundred thousand." "Give me a billion expressions of my ideas." Expressions are commodities. All that matters is the ideas.

By the way: ideas? Uncopyrightable. Machine-created expressions of those ideas? Uncopyrightable. It's uncopyrightable all the way down.

Bill Gates said he's been through three revolutions in his career. He helped enable the computer revolution, where there was zero marginal cost to *duplicate* things; that was in the 1980s. In the 2000s, there was zero marginal cost to *distribute* things through Amazon shopping or email. And now there is zero marginal cost to be able to *ideate* and *create* things.

This is the first time that idea and expression, they've merged. Ideation? Uncopyrightable. Expression? Uncopyrightable. Where does the workforce go at this point? Ideas are the thing, not expressions.

One of the funniest and also most profound cartoons I've ever seen shows a guy saying, "Hey, look! I took this bullet point, turned it into an email I pretended I wrote." And then the recipient says, "Look, I took this email, turned it into a bullet point that I pretended I read."

Funny, but if it *started* with a bullet point (idea) and *ended* with a bullet point (idea), what's the point of this email (expression)? You can make 1,000 versions (expressions). You can make a billion versions (expressions) of this. The ideas are the thing. The facts are the thing. The ideas and the facts: uncopyrightable. Machine-created stuff in the middle? Uncopyrightable. It's uncopyrightable all the way through.

Ideas are valuable. I was spending every waking moment thinking about those three pages of ideas. Those are valuable. The expression that "version one" came up with—or the "version one billion" came up with—it's a commodity.

Ideas are uncopyrightable under the idea-expression distinction. Facts are uncopyrightable. The expressions are copyrightable if there is a modicum of creativity. I've argued in my *All the Music Project* that there is no modicum of creativity, and, again, the U.S. Patent and Trademark Office or the U.S. Copyright Office essentially said, "If machine-created, then uncopyrightable."²⁵

I have a friend named Michael Bommarito. He's one of the guys who beat the bar exam: with GPT-4, he beat ninety percent of humans on the bar exam. He took the *Federal Register*—which is good bedtime reading, if you've ever read it—and he told the large language model, "Express today's *Federal Register* like a chill

²⁵ Thaler v. Perlmutter, 2023 WL 5333236, at *6 (D.D.C. Aug. 18, 2023).

pirate lawyer.” And it took the ideas and facts and said, “*Sorry to disrupt your morning tide, but the compliance crew conducted a sunset review, and fear ye not, the Tariff Act was all above board, those pesky regulations, wishing you a calm sea and well-deserved booty. Signed, your super chill pirate lawyer.*”

Funny, right? But you could also say, “Explain it to me like I’m a six-year-old.” “Explain it to me like my client, who is a high school dropout.” “Explain it to me like someone who has a Ph.D. in physics.” Those expressions are all doable right now, and they’re commoditized. And by the way, those expressions? All uncopyrightable.

Look at the Gettysburg Address. Then, think of the Gettysburg Address as ideas in bullet points. Which is easier to read and understand? And which is poetry? The bulletpoint form is not poetry, but can you understand what he’s saying more quickly? *Hell yeah.*

Here’s one of Justice Oliver Wendell Holmes Jr.’s legal opinions. Here’s Justice Holmes’s ideas, not in an eighteenth-century expression that really takes you all a long time to think through, but in easily digestible bullet-point form. Does this look like a law school summary? *Hell yeah.* Which is easier to skim and read? Which is easier to prep for your bar exam?

This is the thing. The *ideas* are the thing. The expressions—the eighteenth-century expression of what the copyright law is, or the explaining it for a sixth grader—are commoditized.

My wife is an English professor. Since ChatGPT came out, she’s said, “I want to retire. ChatGPT does A work on every assignment. What am I even doing?” And I responded, “You thought that you were teaching writing, but really what you’re teaching is ‘idea transfer.’”

Because what am I doing right now? I’m trying to get my ideas into your brain. I’m doing it through the air and the waves that are going through it. Maybe some of my ideas are going to land in your brain to make you have ideas. But we found as a society that maybe it’s better if we put those ideas onto paper. And now I can maybe get the ideas from my brain onto the paper into your brain as you read it.

So, I said to my wife, “You thought you were teaching writing, but really what you were teaching is idea transfer.” Taking ideas and transferring them to paper, to then transfer those ideas into your brains. And maybe we’re doing that, with the large language

models, cheaper and faster than we ever have before. That process of getting ideas onto paper and into the recipient's brain is just faster than it used to be.

This presentation is me trying to get my ideas into your brain—and maybe, hopefully, they'll stay there when you leave.

Canadian philosopher Marshall McLuhan in the 1970s said, "The medium is the message."²⁶ He was talking about books that turned into radio, TV, movies. Though he wasn't around when the internet came out, it similarly applies to the web, email, the cloud, and now LLMs. Of course, the medium is the message all the way down.

In fact, when I was prepping for this Keynote Address, I was trying to remember "Which of the media was McLuhan talking about?" I asked the large language model to tell me which media he was talking about. So, in that way, literally, the medium *is* the message. The *way* that I use this information is as important as the information itself. The large language model is the thing. The medium is the message.

And how do readers read these days? They read in bullet points. They read in summaries. Uncopyrightable ideas. Snippets. Turns out lawyers are 2024 readers. Judges that we appear before are 2024 readers.

I work for a legal tech company that has a billion legal documents. I've got the essence of every judge's opinion that I'm now inputting into a tool to say, "I want my brief to be in this judge's style," because everybody likes the way that they speak, and everybody likes the way that they write. So, judges are 2024 readers, and then they say, "I like the ring of this brief." They won't know that I've essentially copied that judge's style and applied it.

Legal data is structured data. We don't think of it as structured data, but it's got a lot of structure: for example, a motion to dismiss for lack of subject matter jurisdiction. Standards Advancement for the Legal Industry ("SALI") is a legal data standard²⁷ that I'm leading that's being used by Thomson Reuters, Lexis-Nexis, NetDocuments, the biggest corporations in the world like Microsoft, Facebook, and world's largest firms. Everyone is coalescing around the SALI legal data standard.

²⁶ MARSHALL MCLUHAN, UNDERSTANDING MEDIA: THE EXTENSIONS OF MAN 10 (McGraw-Hill, 1964).

²⁷ See Modern Legal Industry Standards, SALI ALLIANCE, <https://www.sali.org/> [<https://perma.cc/KE3X-KV7E>] (last visited Apr. 16, 2024).

What if you were to take all of these things like lack of standing, preemption, Digital Millennium Copyright Act (DMCA), breach of contract, etc., and tag it up? Tag up the spans of the document with each one of these tags?

That's what I'm doing today. I'm tagging up everything that matters in the law with a billion legal documents, not just in the United States, but in a hundred countries worldwide. I'm doing this for the United Kingdom. I'm doing this for Latin America. I'm doing this for Asia. I'm tagging up the things that matter in every single one of them. I'm working with a group right now that is tagging up all the things in India.

SALI is a non-profit.²⁸ I'm a volunteer for the non-profit. Everything I'm showing you is free and open source. Free as in beer, free as in speech. You can make it extensible, and you can get it for free. It's on GitHub. These are the companies that are using it: Thomson Reuters, Lexis, Bloomberg, NetDocuments, Time Manager, and in-house counsel at Microsoft, Intel, and DLA Piper. They're all coalescing around the legal data standard called SALI that I'm leading worldwide.

Once you extract everything that matters from the documents and dockets, you can say, "Hey, that statement in your brief states that "to determine whether common issues predominate...." We've now built a tool that is able to say, "Here are all the cases with that proposition where the *defendant* wins—and here are all the cases with that proposition where the *plaintiff* wins." So, I can say, "Hey, you represent the defendant, but you've cited *Benton*, where your side lost. Why didn't you cite one of these case with that proposition where your side won?" These are all quantifiable things and it's just math. It's just data. Ideas are the things that matter. Increasingly, expressions don't matter.

When you think about vector space, linguistic concepts are ideas plotted out in vector space. There are two dimensions in this graph, and so you can see it takes "shower," "kit," "valve," "garden hose," etc., and the AI kind of clusters them together. This is two-dimensional vector space: X and Y axes. Now imagine adding a third dimension: a Z axis. You could see where these things cluster in three-dimensional vector space? Now imagine adding a fourth dimension. Can't do it, right? It's impossible for our human brains to figure out how to visualize a fourth dimension.

²⁸ *Id.*

But now imagine 12,000 dimensions. That's what the large language models are doing: They're plotting concepts into 12,000 dimensions. And somewhere in those 12,000 dimensions are "Bob Dylan-ness," and "Ernest Hemingway-ness," and "Picasso-ness." The idea of each of those and the style of each of those live in the 12,000-vector space. And so when you vectorize things, you're plotting them into this 12,000-vector space, which is just the ideas of Bob Dylan and the style of Ernest Hemingway.

And can Bob Dylan sue me for writing a song "in the style of Bob Dylan"? No. The style of Bob Dylan is not copyrightable. If it were, he could sue every singer-songwriter since the 1970s, right? The style of Bob Dylan is uncopyrightable because it is an idea.

So then why would a machine extracting his ideas to create a new expression—an expression that is uncopyrightable—why would that be any different than when a *human* does it? The input, if it's just ideas. If it's pulling text from the book, it's pulling those ideas, putting them into vector space, and throwing away the expressions. The expressions go away. All that's left is the idea. The expressions go into the trash. And then those ideas are facts that are unoriginal and uncopyrightable. Then you get new expressions that are also uncopyrightable.

Ideas matter. Expressions don't.

The courts also consider fair use: Are LLMs transformative? Yeah, they are. If you think about the Google Scholar case, the Second Circuit held that if you index a whole bunch of books, and you take that word index, that is a transformative purpose.²⁹ By taking words and being able to search those words, that is transformative—beyond what the books were originally intended for.

Now, if a *word index* is transformative, how about a vector model in a 12,000-dimensional vector space? If an index of words is transformative, how transformative is the ideas that are in that 12,000-dimensional vector space?

So, the fly in my ointment of this argument is: what if you were to take the first twelve paragraphs of *Harry Potter* and say, "Tell me what's statistically likely after that." And then it spits out the entirety of *Harry Potter*. Then the input and the output are identical.

Because, of course, after eight paragraphs—and this is what happened in the *New York Times*: Plaintiffs prompted eight

29 Authors Guild v. Google, Inc., 804 F.3d 202 (2d Cir. 2015).

paragraphs of it and said, “What’s statistically likely after that?”³⁰ And the LLM spit out the rest of the *New York Times* article. Of course it’s going to do that, right?

In cybersecurity, they call that “red teaming.” That’s either bad actors—or people *pretending* to be bad actors—trying to get the machine to do something that it’s not supposed to do. In that case, though, isn’t the *New York Times* being the bad actor, trying to get the machine to do something that it’s not supposed to be doing?

And in those cases, isn’t that kind of like a videocassette recorder (“VCR”) from the 1980s Sony “*Betamax* case,” where the plaintiffs argued “People can use VCRs to be able to infringe copyright.”³¹ In the case, the U.S. Supreme Court focused on whether there are there substantial, non-infringing uses of the VCR—and it turns out, there are. You can record HBO movies, and you can *also* record your kid’s recital. These are substantial, non-infringing uses. So, the VCR was non-infringing.

How many substantial, non-infringing uses are there for a large language model? It’s going to transform the world. You know, a billion, literally a billion substantial, non-infringing uses.

So how does that fall into the *New York Times* argument? The thing they need is guardrails. Make sure that the output doesn’t match the input, and that’s what OpenAI is doing, and that’s what every foundational model should be doing going forward.

Is this painting copyrightable? [Displays abstract art.] A machine made that. A guy named Stephen Thaler tried to get copyright registration for this. The U.S. Copyright Office said, “No. Machine-created, therefore uncopyrightable.” Can machines create patents? No, machine-created, so unpatentable.

That makes sense, right? Because really, what is intellectual property but the government saying to you, “I’ll give you a limited monopoly to incentivize you to make more things. Essentially, this is a quid pro quo: I want to incentivize you to make more things. So, I’m going to give you a monopoly.”

Machines don’t need incentives. They don’t need an incentive to do make many versions of my article. My server didn’t need an incentive to make 471 billion melodies. It literally took four days to create 471 billion melodies. There’s no need to incentivize for

³⁰ See Aatish Bhatia, *Watch an A.I. Learn to Write by Reading Nothing but ‘Harry Potter’*, THE N.Y. TIMES (Apr. 27, 2023), <https://www.nytimes.com/interactive/2023/04/26/upshot/gpt-from-scratch.html> [https://perma.cc/7QC5-L7EY].

³¹ Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417 (1984).

that. So, should my melodies be copyrighted? No, because the machine needs no incentives. The incentives are off.

There is a case about *Zarya of the Dawn*, where a graphic novel author filed their work with the U.S. Copyright Office, which rubberstamped it. Then the author said, “Wow, I didn’t think they would register it, because all the images were made by AI [Midjourney].” And the U.S. Copyright Office said, “Huh?” And the U.S. Copyright Office revoked the copyright registration, saying, “Alright, you get copyright for what you wrote as a human, but the machine-created things: uncopyrightable.”³²

So, they split it: If a human created it, it’s okay and copyrightable, but the machine-created parts, uncopyrightable. So now if you file something with the U.S. Copyright Office, you have to declare how much of it was human-created and how much was machine-created.

But how about the my article I mentioned earlier, where I took my ideas (three pages’ worth), the machine spit something out, and I spent three hours jamming with it—editing, revising, and putting it into my voice?

This is back in 2020 before large language models. I used a tool called These Lyrics Don’t Exist. I said, “Give me a country song that’s a love song and happy,” and it spit out a bunch of lyrics called “Freight Train of Love.” And then I created this song that I put up on SoundCloud with the melodies from the *All the Music Project*, lyrics from “These Lyrics Do Not Exist” (an AI songwriter), drums from “Superior Drummer” (an automated drummer), and then I arranged and did other things. How can I unbake the cake as to what is copyrightable and what is uncopyrightable in this? I did lots of things. What’s copyrightable?

Really, we’ve been dealing with this dilemma for hundreds of years with joint works. If book coauthors’ contributions are 90% to 10%, do the courts decide “John gets 90% of royalties, Jane gets 10%”? No, courts aren’t going to settle fights about, “is that 10% the *more important* part? What are the splits going to be?” No. Everybody owns an *undivided interest* in the work’s entirety. We’re not going to unbake the cake, the courts say.

³² See U.S. Copyright Office limits registration for AI-created *Zarya of the Dawn* graphic novel, NIXON PEABODY (Mar. 9, 2023), <https://www.nixonpeabody.com/insights/alerts/2023/03/09/us-copyright-office-limits-registration-for-ai-created-zarya-of-the-dawn-graphic-novel> [<https://perma.cc/8VKR-P63Z>].

How about co-authoring with machines? When I wrote my article by jamming with the machine, I contributed a bunch, then it contributed a bunch, and then I contributed a bunch more. Human, then machine, then human. How do you unbake that cake? And the answer is: you don't because you can't. It's impossible.

Because if you look at a redlined Track Changes sentence, are you going to say, "Okay, the black stuff that the machine created, that's uncopyrightable, but I get the two words here and there that I wrote and edited: those two words are copyrightable." If humans were doing this as a joint work: undivided interest, right? Because I'm not going to be able to unbake that cake.

If I tried to register the article, how would I tell how much was human and how much was machine? Because, honestly, I didn't write it with "track changes." I have no idea what was the human's and what was the machine's. I jammed with my machine coauthor. So what parts are machine-created? What parts are human created? I don't know. It's one whole, undivided interest.

Music is even worse. Music doesn't have "track changes." I can ask a machine "Give me a bunch of chord progressions." Machine spits out a bunch of chord progressions. Give me a bunch of melodies? Machine spits out a bunch of melodies. I'm up till 3 a.m. jamming with that, moving things around, editing here and there, then getting out the door. How do you unbake that cake? You don't. Because it's impossible. Undivided interest whether it's a human or machine. You can't unbake cakes.

So how do I comply with the U.S. Copyright Office, saying how much was human and how much was machine if I don't even know myself? Am I committing fraud on the U.S. Copyright Office if I say I don't know?

And how many of these things are filed with the Copyright Office without that person even saying anything? How many hit songs have actually been in this hybrid process where the human and the machines are jamming, and they don't disclose it? And you're never going to find out until it comes to a lawsuit. And even if it goes to a lawsuit, you're not going to find out until you're in the deposition. And even in the deposition, you're not going to find out if the person lies and says, "No, that was all me." So, the odds of catching this kind of fraud is next to zero. Where does this land? It's *uncopyrightable turtles* all the way down.

Old music is also eating new music. Songwriters are selling their catalogs: Bruce Springsteen,³³ Sting,³⁴ Bob Dylan,³⁵ Leonard Cohen.³⁶ Everyone's selling their catalogs because they can make a lot of money, hundreds of millions of dollars, on these things.

When we think about the Old World—maybe before 2000—new music was the most-important thing. Record labels invested in artist development. They spent millions to make sure that new artists get airplay. Artists and repertoire (A&R) was a thing.

Catalog music, though? You go to the gas station and get old (catalog) music for a dollar. Nobody cared about catalog music, right?

But then that flipped. Going forward, record labels dropped their A&R budget. They're not developing people that blow up on TikTok. They're like, "Cool, you blew up on TikTok. Now we know you're viable, but I'm not going to invest a dime in new artists these days." Catalog, though? Bob Dylan, Springsteen, Sting? The industry is investing billions in that. This is old music eating new music.

And I was actually at South by Southwest with the two lawyers that do a lot of these deals with Bob Dylan, Leonard Cohen, etc., and I asked them, "Is this an example of the old music eating new music?" And they responded, "Yeah, of course it is."

Because, if you're an investor, which do you bet on? (1) Bob Dylan's songs' ability make money? Or (2) Someone out of South Los Angeles, whose songs have a one-in-a-thousand chance of making money? If I'm going to invest my money, will I do it on Bob Dylan or on this unknown artist? Of course, I'm going to invest in Bob Dylan. Old music eats new music.

Also "you stole my melody" lawsuits: old music eating new music. Whether it's Tom Petty suing Sam Smith, whether it's The Hollies suing Radiohead suing Lana Del Rey, whether it's the "Blurred Lines" case. This is old music eating new music, or at least trying to.

³³ Ben Sisario, *Bruce Springsteen Sells Music Catalog in Massive Deal*, The N.Y. Times (Dec. 15, 2021), <https://www.nytimes.com/2021/12/15/arts/music/bruce-springsteen-sells-music-catalog.html> [<https://perma.cc/7AX8-UBZZ>].

³⁴ Ben Sisario, *Sting Sells His Songwriting Catalog for an Estimated \$300 Million*, The N.Y. Times (Feb. 10, 2022), <https://www.nytimes.com/2022/02/10/arts/music/sting-sells-catalog-universal.html> [<https://perma.cc/D6JG-CAKS>].

³⁵ Ben Sisario, *Bob Dylan Sells His Songwriting Catalog in Blockbuster Deal*, The N.Y. Times (Dec. 7, 2020), <https://www.nytimes.com/2020/12/07/arts/music/bob-dylan-universal-music.html> [<https://perma.cc/TLU2-XRN2>].

³⁶ Jem Asward, *Leonard Cohen Song Catalog Acquired by Hipgnosis*, VARIETY (Mar. 5, 2022, 6:25 PM), <https://variety.com/2022/music/news/leonard-cohen-hipgnosis-1235196842/> [<https://perma.cc/2R49-DBC4>].

My buddy Michael Bommarito, one of the guys who beat the bar exam, said, “I really like your copyright thing. Let’s do the same thing with patents. Let’s take every single patent that has ever been filed and take every claim in each of those patents, and then use a large language model to recombine all those claims into new inventions—new prior art.” So, we’re going to upload all those new inventions, that recombined prior art, into the Internet Archive so that if you try to recombine old claims into a new patent application, the Patent Office and the Examiner will say, “No, Damien, Mike, Noah, and All the Patents did that in 2024. You can’t get a patent monopoly because that’s already been invented. As prior art, you don’t get a patent for it.” So, really, patents should be for whatever inventors do that is novel; not just recombining old things. So, what we’ve done for copyright, we want to do for patents as well.

This is a patent application that has my name on it.³⁷ It’s by Thomson Reuters, and the application is in the public domain, so that’s why I can talk to you about it. The patent is about workflow and practice management, and it’s essentially saying: if you have a matter—whether it’s litigation or transactional—you can break that matter into 1,000 tasks.

Then for Task 357, I, as a partner, can assign it to Jane. Jane can now see the ways to complete Task 357. Here are the Westlaw queries to do for Task 357. Here are the internal documents that have done Task 357. And here are your colleagues who have done Task 357 in the past, if you want to talk with them. Do that for Task 357—and all thousand tasks.

Jane then moves Task 357 from “To Do,” to “Doing,” to “Done.” Now you firms know how long it takes to do Task 357. And because you know how long it takes, you know how much it costs to do task 357. Now that you know how much task 357 costs, when new things come in the door, you can budget more accurately because now you know how much task 357 and all the other things cost.

The last thing I’m going to talk about is human creativity. What is creativity but statistical unlikelihood? Sting said, “The essence of all music is surprise.”³⁸ He essentially said, “If I’m not surprised in the first thirty seconds, I go to the next song.”

³⁷ Systems and Methods for Workflow and Practice Management, U.S. Patent No. 11868936 (filed May 23, 2017) (issued Jan. 9, 2024).

³⁸ See *Interview: Sting Talks Tour & New Album 57th & 9th*, TICKETMASTER (Nov. 14, 2016), <https://blog.ticketmaster.com/interview-sting-new-album-57th-and-9th/> [https://perma.cc/UL8Y-4HWY].

“Surprise” means it’s “statistically unlikely.” Ray Bradbury essentially said, “Creativity is continual surprise.”³⁹ Surprise: statistically unlikely. Joseph Conrad essentially said, “It’s an education of the unseen, familiar, and surprising.”⁴⁰ Statistical unlikelihood. Thurgood Marshall’s arguments in the 1950s: statistically unlikely. Who’s going to be making those arguments? Only a few people are going to do that. Legal creativity is statistically unlikely.

I’m applying that to legal tech. I took the *New York Times* case, where they sued Microsoft and OpenAI just a few weeks ago,⁴¹ and I dragged the Complaint into my tool that I’m building right now, where we extract every single claim and every fact—everything that I, as a litigator, cared about, when I got a complaint. What are the claims? What are the facts? What are they looking for? What is the relief?

Then we go into strategy. What are the defenses to each of these claims? Give me a questionnaire I can ask my client to bolster my answer to these defenses. Find similar things in 800 million dockets and documents and motions, briefs, pleadings, motions that have happened, cases like mine.

Then, the output is legal questions specifically focused on this complaint. What are the legal standards for copyright infringement in the context of GenAI and large language models? How have the courts interpreted fair use for copyrighted material for training AI systems? You upload a document; it gets these things out of the box.

That’s why I was in New York yesterday. Tonight, right after this, I’m going to speak to the American Bar Association (ABA) about what I’m doing at vLex. For each of the claims, it extracts all the citations, governing law, and related facts for each of the claims. What are the defenses? “Hey, you might want to argue fair use, that training is transformative and doesn’t substitute for the

³⁹ See QUOTEFANCY, <https://quotefancy.com/quote/897626/Ray-Bradbury-Surprise-is-where-creativity-comes-in#:~:text=Ray%20Bradbury%20Quote%3A%20%20%20%20where%20creativity%20comes%20in.> [https://perma.cc/8AEC-NHQB] (last visited April. 16, 2024).

⁴⁰ See Maria Popova, *Creative Magic and What Makes a Great Writer: Joseph Conrad’s Beautiful Tribute to Henry James*, THE MARGINALIAN, <https://www.themarginalian.org/2015/12/03/joseph-conrad-henry-james-appreciation/> [https://perma.cc/2ZQ2-XHRQ].

⁴¹ See Halelua Hadero and David Bauder, *The New York Times Sues OpenAI and Microsoft for Using Its Stories to Train Chatbots*, ASSOCIATED PRESS (Dec. 27, 2023), <https://apnews.com/article/nyt-new-york-times-openai-microsoft-6ea53a8ad3efa06ee4643b697df0ba57> [https://perma.cc/P5DX-EZRH].

original.” This is all the machine spitting these things out. “Hey, what do I need to show? That the large language model is educational and research-oriented, and it doesn’t affect the market for the *New York Times*.” That’s a good legal observation.

It did that for every single one of them, and I asked for one defense, but I could have asked for 100 defenses or 300 defenses. And it will give me 300 defenses that I can pick and choose from the best ones. Give me a questionnaire to ask my clients. “Hey, how did you create those datasets?” “How did you select what you were going to do?” “How much of the *New York Times* stuff did you pull?” “How do you try to avoid copyright?” “How many lawyer hours were spent thinking about questions like this that my machine literally spit out in less than a minute?”

Let’s talk about the idea-expression dichotomy. Training the large language model is merely extracting the ideas that are vector embeddings, and not the expressions, which are jettisoned. The output is something you can paste into a brief: “Your Honor, here’s the idea-expression dichotomy” and “Your Honor, here’s how large language models are trained.”

What I didn’t show you is the rest of the brief that you can copy and paste and put into the output; the vector embeddings of our ideas; how we throw away expressive content; and the legal precedent. We do that, not just for a complaint to an answer, but we’re building every single motion, brief, pleading, deposition, etc. Bring your own facts. Go ahead and record your interview with your clients, transcribe that interview, upload that into my system. You got a ready-to-bake complaint, ready-to-make answer, and ready-to-make motion all the way through.

I’ve spoken to Fortune 50 companies. What they want to do is take every complaint that’s ever been received and compare that historically to what their law firm has done, seeing if the lawyers added anything on top of what the machine did. And if their lawyers didn’t do anything on top of that, that says something.

Then, going forward, before the corporations even give the complaint to their law firm, they’re going to run the Complaint through this tool. And they’re going to give the machine’s output to their lawyers, saying, “What can you do on top of what the machine spit out in a minute? Because that’s all we’re going to pay you for. We’re not paying for you to re-create the wheel.” They’re going to do that for every single one of the complaints.

So, this is kind of dire—you being law students—right? But this is the question—these are the honest things that people say, like, “ChatGPT, it’s not quite there yet.” They don’t see the train a-comin’ down the track. I’m telling you what’s coming: it’s not coming down the pike. This is out in the world now. This is not the future; this is today.

So, as law students, you should think about what you can do that’s on top of what my machine can do today in 2024 because my machine is going to do more in 2025 and more in 2026 and more in 2027, and inevitably where this lands is our humanness—our humanity. I use TurboTax every year. I still go to a human accountant to ask, “You know, tell me about this. I’m not quite satisfied. Talk me off the edge here, is this right?”

Our humanity is what we give. We are counselors. So, really, the more we can counsel, the more we’re going to make it through our large language model.

Depending on the day, I vacillate between: the legal profession will go away—that’s a scarcity mindset—or an abundance mindset, where my friend who serves as litigation counsel at Ford Motor Company says, “Do you know how much we’re regulated and how much of that regulatory work we’re not giving to you because you are all too expensive? If you use large language models and shrink your costs, I’m going to give you way more work.” So that’s an abundance mindset.

So, depending on the day, I vacillate between scarcity and abundance. Is all the legal work going to go away, or are we going to have more work than we ever have? Today, I’m focused on abundance.

Think about the access to justice problem: eighty percent of legal needs are unmet because we’re too expensive. What if we were to shrink the cost and open up that eighty percent to help not ten people—because of scarcity—but abundance: I can now serve 1,000 people. I can now work for a legal aid organization that can serve 10,000 people. That abundance is going to make society way better. And it’s *because* of the machines, not *in spite of* the machines.

Large language models produce output that’s statistically likely. They tell you the statistical likely next word, sentence, etc. That is uncreative. Humans are jagged. Less predictable. The way that ChatGPT detectors are saying “if it’s jagged, it’s more likely to be human created.” That’s what they do.

But what if you were to *force* unlikelyhood? What if you were to tell the LLM, “Give me gangster rap in the style of the *King James Bible*.” The LLM writes: “*In land of the hustle brethren gather round / Behold the tale of the streets profound / In the beginning God created Heaven and Earth / But in the alleyways, man proves his worth. . .*”

Those are two statistically unlikely things that are creative, right? If a human wrote that: “Oh, that person’s creative,” right? That’s the machine doing that.

The idea of gangster rap and the idea of *King James Bible*, those each live in different vector spaces. And what the LLM is doing is finding the connective *latent* space—that’s the term that’s used—latent space between these two vectors.

Then I said, “You’re a creative poet. Create ten poem ideas, creating disparate concepts that usually aren’t associated with one another.” The LLM combined things like “ancient ruins and modern technology”; “classical music and urban street decay”; “chess strategies and beekeeping.”

This one seemed interesting to me: “mushroom foraging and cybersecurity.” Those are two very disparate concepts. So, I said, “Prompt something conveying mushrooms and cybersecurity,” and this is its output:

*Here in the forest of code, mushrooms hide
Data encrypted in nature’s stride
A forager’s keen eye, a hacker’s mind
In each cap and byte, secrets intertwined.*

That’s creative, right? It’s statistically unlikely. That’s a machine being creative. So, really, maybe the machine could say, “Hey, figure out what is far away in latent space and then shoot right in the middle and give me connections between those two disparate things.” Do that for poems. Do that for legal arguments. Facts. Law. Give me a creative legal argument that is going to win for this judge, and statistically likely to win for that judge. So, it’s the cosine distance—people who don’t like geometry, fine, but it’s kind of the middle.

You don’t want to have the cosine similarity be too close together because that’s obvious; that’s not creative. But you also don’t want to have it too far away, like that’s super weird, right? And I don’t want to do that. But there’s a creative cosine to say, “Yeah, it’s not too super weird, and it’s not too on-the-nose.” There’s a middle

distance to be able to say, “That is creative.” Just enough to be creative. To say, “Oh, that’s so clever.” The LLMs are doing this.

I’m going to close with a love poem, comparing a human relationship to assembling an IKEA furniture set. And it’s hilarious, right? Who would have thought to put love with an IKEA furniture set? But the output is:

*Like love, this bookcase comes in pieces
A puzzle of boards and screws.
We fumble with vague instructions
Baffled, but determined
Trying to fit tab A into slot B.
It tests our patience, teamwork
Whether we can weather the frustration
And see the vision for the finished whole.
We stand back to the end with sore thumbs
And look upon our creaky creation.*

That’s creative. It’s stupid that I got emotional, but it’s—I was thinking about my wife, right? That’s beautiful. A machine created that.

We’re in a weird world—in our law and in our creativity. What is “creative” these days? So, I would ask everybody in this room: think about where you fit into our new world that I’ve been talking about with you, because it’s not tomorrow; it’s today. And anybody who tells you that we’re not going to be doing anything with large language models in law anytime soon doesn’t see the train coming down the tunnel. This is happening.