# INTERVIEW OF THERESA WEISENBERGER[*] AND CHAD RUTKOWSKI[**]

*Conducted by* ELIZABETH TOWNSEND GARD[***]

*Chad Rutkowski and Theresa Weisenberger, both attorneys at BakerHostetler, are two leading voices in the conversations surrounding copyright and artificial intelligence. Elizabeth sat down in April 2025 for a conversation on how we might think about our current state.*

**Elizabeth Townsend Gard:** Thank you so much for joining me. You are both so important to the current AI conversation. It's a thrill to have you here to chat. I really appreciate you being part of this.

**Theresa Weisenberger:** Thank you.

**Chad Rutkowski:** We appreciate being asked that Elizabeth? Obviously, it's near and dear to our heart. And I tell everybody that listens. You know the Copyright Society made me a copyright lawyer. but for the Copyright Society I wouldn't be doing any of this stuff.

**Elizabeth Townsend Gard:** Awesome. Okay? So I'm a historian, an oral historian by training. So we're going to start with a really simple thing which is for each of you to tell me your name, the date, and where you're calling from, and then I'm going to ask you each for a little biographical information again for preservation purposes. So, Theresa, should we start with you.

**Theresa Weisenberger:** So, Theresa Weisenberger, I'm calling from Atlanta, Georgia. I really wanted to say, like "longtime listener first time caller."

---

[*] Theresa Weisenberger is a Partner at BakerHostetler where she co-leads the Artifical Intelligence practice. She focuses her work in the areas of intellectual property and technology law, particularly in counseling on emerging technologies and its intersection with the evolving legal landscape. Theresa also serves as the Secretary and Vice President of Governance of the Copyright Society. *Theresa Weisenberger*, BAKERHOSTETLER, https://www.bakerlaw.com/professionals/theresa-m-weisenberger/.

[**] Chad Rutkoski is a Partner at BakerHostetler where he co-leads the Technology and IP Transactions and Outsourcing team. He is an expert on the intersection of copyright and technology, especially on copyright issues with AI and other softwares, and is the author of Lexis Practical Guidence notes on data licensing and Creative Commons licenses. Chad currently serves as the Chair of the Copyright Society's AI Subcommittee and previously developed the curriculum for the Copyright Society's AI programming series. *Chad Rutkowski*, BAKERHOSTETLER, https://www.bakerlaw.com/professionals/chad-a-rutkowski/.

[***] Editor-in-Chief, Journal of the Copyright Society, and John E. Koerner Endowed Professor of Law, Tulane University Law School

I am a partner at BakerHostetler, where I co-lead the AI practice and our IP transactions group.[1] I have a background in electrical engineering, and early in my career I focused almost exclusively on patent law.[2] I address copyright and AI from a technology perspective. I'm particularly interested in how copyright law both protects and restricts the development of technology, and AI is an ideal case study. As part of my focus on AI, I maintain the AI Copyright Case Tracker where we monitor how copyright law is being applied in the context of the development and use of AI technology.[3] And I think that's it.[4]

**Chad Rutkowski:** So I'm Chad Rakowski. I'm a partner in the Philadelphia office of BakerHosteler.[5] My practice focuses on the intersection of copyright and technology. I write rather regularly on the copyright protections for data and copyright protections for software, and how that all relates and interrelates with AI. I've been writing about AI issues since 2016, when the Copyright Office issued its the Third Edition of its Compendium of U.S. Copyright Office Practices.[6] The Copyright Office was updating guidance on authorship issues, including AI authorship.[7] And I've been giving presentations on AI related topics

[1] BakerHostetler's Artifical Intelligence practice has provided a consortium of more than 40 companies with AI advice in a variety of areas, such as data collection, development of AI policies, and data privacy. The AI practice focuses on three main concerns: the use of AI in the legal field, the use of AI in the work and operations of clients, and the emerging legal and regulatory requirements for the use of AI. *Artificial Intelligence (AI)*, BAKERHOSTETLER, https://www.bakerlaw.com/services/artificial-intelligence-ai/.

[2] Theresa received her B.S. in Electrical Engineering from Mississippi State University in 2008 where she graduated *magna cum laude* and with a certificate in Engineering Entrepreneurship. *Theresa Weisenberger*, *supra* note 1.

[3] The AI Copyright Case Tracker monitors key pending litigation in the U.S. that raise copyright issues related to the creation or use of generative AI. Each case within the tracker includes an overview of the lawsuit, key filings that address the substantive legal issues, and a summary of the filings. Theresa M. Weisenberger, *Case Tracker: Artifical Intelligence, Copyrights and Class* Actions, BAKERHOSTETLER, https://www.bakerlaw.com/aicasetracker.

[4] Theresa is also the current Secretary/VP of Governance for the Copyright Society.

[5] *See Chad Rutkowski*, *supra* note 2.

[6] Chad's first article about AI discussed section 313.2 of the *Compendium of the U.S. Copyright Office Practices, Third Edition*. The article explored how AI technologies that can produce creative works challenge traditional ideas of authorship and ownership, as well as what the legal considerations are for protecting the copyright of AI-generated works. The full article can be found here: Chad A. Rutkowski, *The Monkey in the Machine*, BAKERHOSTETLER (Feb. 3, 2016), https://www.ipintelligencereport.com/blogs/the-monkey-in-the-machine/.

[7] Section 313.2 of *Compendium of the U.S. Copyright Office Practices, Third Edition* discussed Works That Lack Human Authorship, such as those produced by a machine. In that section the Copyright Office declared that it would not register machine created works if there was no creative input or innovation from a human author. U.S. COPYRIGHT OFFICE, COMPENDIUM OF U.S. COPYRIGHT OFFICE PRACTICES § 101 (3d ed. 2021), https://www.copyright.gov/comp3/docs/compendium.pdf.

and data ownership and copyright fair use for the Copyright Society and other organizations since 2018.[8] I'm the co-chair of the firm's Intellectual Property and Transactions Group, and co-chair of its AI group.[9] And I am the current chair of the Copyright Society's AI Programming Committee, and the co-chair of the American Bar Association, Copyright and Tech Committee, as well.

**Elizabeth Townsend Gard:** So do you just dream of AI issues like, do you ever get away from it?

**Chad Rutkowski:** For many years I felt like some of the disturbed folks that you might find in front of a 7-Eleven, trying to convince everybody that the end was coming. And I think the kindest compliment that I've gotten over the past two years was from our fellow Copyright society member and New York Chapter chair. James Trigg,[10] who is at Kilpatrick, who told me at a dinner in 2023 after our first Copyright Society AI Series panel, that for years he would listen to my presentations on data, copyright, and machine learning, and he would say to himself, "What is he talking about? And why do we care about this?" … And you know the compliment was that I was right, that this was something people needed to be paying attention to. And so here we are.

**Elizabeth Townsend Gard**: So help me help me understand, Chad, why did you start "paying attention" in 2016? What made you think about it in 2,016?

**Chad Rutkowski**: So in 2016 my practice was changing in a lot of different ways. I had been doing quite a lot of work in the patent space, although I don't have a technical degree. So I was getting a lot of exposure to issues around software architecture, software, coding, and software networking. And so I now had this sort of buildup of knowledge and my work in the patent space was beginning to diminish. So I had always had this interest in copyright law and decided that I wanted to figure out how to join those two things together, especially because nobody was really interested in the copyright space in tackling issues around code and data ownership. It was a bit of a backwater. I was focusing at the time on big data analytics, which was the buzzword.[11] You know the

[8] In 2020, Chad was a speaker for the Copyright Society's Content as Data and Data as Content program, where he spoke about how AI and machine learning interacts with copyright. *Content as Data and Data as Content: Copyright Issues in Machine* Learning, THE COPYRIGHT SOCIETY, https://copyrightsociety.org/event/content-as-data-and-data-as-content-copyright-issues-in-machine-learning/.

[9] *See Chad Rutkowski*, *supra* note 2.

[10] James Trigg is a Partner at Kilpatrick where he represents creators and brand oweners in the areas of copyright and trademake law. He has previously taught trademark and entertainment law as an Adjunct Professor at the Univeristy of Georgia School of Law. *James A. Trigg*, KILPATRICK TOWNSEND & STOCKTON, https://ktslaw.com/en/People/T/TriggJamesA.

[11] Big data analytics deals with extractig insights from large amounts of data and complex data sets by systematicly processing and analyzing the data. Tim Mucci and Cole Stryker, *What is Big Data Analytics?*, IBM, https://www.ibm.com/think/topics/big-data-analytics.

Gartner Hype cycle,[12] the issue *de jour* at the time. Related to that, I was beginning to learn more of the connection between data analytics and at least machine learning, a form of artificial intelligence. And so the ownership of data and the copyright issues that go around that were very much being discussed in that context back then, and that includes use of copyrighted content to train AI data models. We think of this as coming to the forefront with the prominence of ChatGPT at the end of 2022 and the beginning of 2023, but realistically, that was being discussed even back in 2016. So I was paying more attention to the AI space at the time. The hot technology was Google's Deep Dream, where you would take a photograph and plug it into the Google interface, and it would use the images that it was trained on to detect things like animals.[13] For example, is this a cat or is this a dog? And it would intentionally scramble the color palette of the photograph and map it to color palettes of animals that were in its image database and sort of create these funky psychedelic type outputs. And: that was sort of novelty art, there was no real apparent value in it, but it raised really interesting questions of authorship. Is there and author who is the author? Is it that you have selection authorship in picking the photograph that you sort of fed into the model or is Google the author because the software does the lion's share of the work? And so, as I began to think about those issues, the Copyright Office issued the Third Edition of the Compendium. The Compendium stated that works created by animals (at the time the Naruto monkey case was all the rage) and the Copyright Office sort of addressed that tangentially in the Compendium and saying, for example, a photograph taken by a monkey is not copyrightable.[14] And then in the same sentence–I think it was comma–works created solely by machines are not copyrightable, and that I found to be found to be bothersome right that the technology was likely to develop in ways we hadn't foreseen yet, that it was going to become an important technology, and to compare what AI was

[12] The Gartner Hype Cycle is a graphical representation of the maturity, adoption, and social application of specific technologies over time. The illustration helps organizations understand the potential value of emerging technologies within the context of their industry. The Cycle has five distinct phases: Technology Trigger, Peak of Inflated Expectations, Trough of Disillusionment, Slope of Enlightenment, and Plateau of Productivity. *Gartner Hype Cycle*, GARTNER, https://www.gartner.com/en/research/methodologies/gartner-hype-cycle.

[13] Deep Dream is a computer vision program created by Google that alters images to create surreal, dreamlike visuals. The program was designed to understand how neaural networks perceive and process images, and it was trained on millions of photos to classify and recognize patterns. Deep Dream was about to use its training to interpret and manipulate images to amplify patterns that would result in otherworldly images. *Deep Dream: An In-Depth Exploration*, GEEKSFORGEEKS (Aug. 1, 2024), https://www.geeksforgeeks.org/deep-dream-an-in-depth-exploration/.

[14] Section 313.2 of *Compendium of the U.S. Copyright Office Practices, Third Edition* declared that works that did not have a human author could not be copyrightable and specifically gave the example that a photograph taken by a monkey could not be protected. *See supra* note 10.

likely to do to what a monkey did was not likely to capture the nuances. So I wrote a blog post cleverly titled the "Monkey in the Machine."[15] sort of teased out some of these issues and some of the scholarship, that was out there at the time. Annemarie Bridy, who was a Stanford professor at the time, who is now heading copyright at Google,[16] was addressing some of these issues, and I relied quite heavily on some of her scholarship and thought leadership.[17]

**Elizabeth Townsend Gard:** And Theresa, how did you get involved and thinking about AI?

**Theresa Weisenberger:** Like I said earlier, I have a tech background, and I started my career doing patent prosecution and patent litigation. My industry focus was primarily software with the Supreme Court decision in *Alice v. CLS Bank,*[18] the patentability of software was really pulled back. A lot of our clients were left trying to figure out how to protect their investment in software.

Around 2016, I was at an outside counsel conference for one of our clients, and I had just started working at the firm, and I get cornered by this guy, Chad, that starts proselytizing about how copyright is the solution to all of our problems and that it's going to be what protects software going forward. And he had this great analogy explaining how the non-literal elements of software are protected just like plot and characters. He talked about Harry Potter. I just started at this firm as an associate, and I got this partner overly excited about copyright law in a room full of patent attorneys. That scared me off for a little bit.

But over timeI heard him talk over and over about copyright and its value for software. When I was hanging out with patent attorneys, when clients asked "How do we protect our software?" They didn't have a good answer. Copyright protection was often dismissed out of hand based on the misconception that infringement requires that source code be copied verbatim–which reminded me of Chad's Harry Potter analogy,. And so I dug a little bit further into this area. And that's how I started working more closely with Chad.

My AI practice began to grow around the same time, both as I was drafting patents to protect machine learning algorithms and the Patent Office was reluctant to issue any software patents.

I did a secondment at a company that had developed technology for running its operations, including AI technology for identifying contaminants in their supply and for identifying potential real estate acquisitions based off of publicly

---

[15] *See* Rutkowski, *supra* note 9.

[16] Annemarie Bridy is Senior Copyright Counsel at Google and a Nonresidential Fellow at the Stanford Law School Center for Internet and Society. Annemarie is the former Allan G. Shepard Professor of Law at the Univeristy of Idaho, having taught courses relating to intellectual property and interformation law. *Annemarie Bridy*, YALE LAW SCHOOL, https://law.yale.edu/annemarie-bridy.

[17] *See generally* Annemarie Bridy, *Coding Creativity: Copyright and the Artificially Intelligent Author*. 2012 STAN. TECH. L. REV. 5 (2012).

[18] Alice Corp. v. CLS Bank Int'l, 573 U.S. 208 (2014).

available data. And what was interesting about that work is that it came right at that time where we were kind of trying to figure out how to advise clients on patenting versus copyright. So I had to dig really far into actually at the specifics of how the technology worked, while I have an engineering degree, I did not specialize in AI. Back when I was in school, I don't think you could specialize in AI. That was kind of my segue into the AI stuff.

And then, to be honest, I got more involved in copyright because of the Copyright Society, and it's not exactly the same story as Chad's. It's not that I became a copyright attorney because of the Copyright Society. I became a copyright attorney because I wanted to work with the people in the Copyright Society. It was such a fun, intelligent organization that delves into really squishy problems, and once I overcame that learning curve, it became a bigger part of my focus.

**Elizabeth Townsend Gard:** Love it So how would you describe the moment we're at? How have we progressed? How do you describe who we are at the moment?

**Theresa Weisenberger:** Before 2022, we already had AI technology integrated into our systems; we just weren't calling it "AI." It wasn't as interesting to talk about data analytics or the machine learning algorithms. But once 2022 hit and AI became the hot topic, largely in part to the public reaction to ChatGPT. There in AI exploded, , and everybody wanted to integrate AI into everything, sometimes regardless of whether AI was the best solution to the particular business challenge. And from the legal perspective, from our clients, we were getting more questions about the risks of AI in general. The tech industry and the government both struggled with not only how to identify those risks, but how to allocate liability and responsibility appropriately. We are all still trying to figure out the right kind of language to use to discuss AI–there is no single definition of "AI" or even "generative AI" or "large language model" that provides enough clarity to regulate this technology. What we've seen in the EU AI Act,[19] in some of the state laws is that they're encompassing more than generative AI–whether intentionally or not. So that while ChatGPT brought to the forefront generative AI and the issues that come up with that,[20] we're now seeing more scrutiny and more regulation of non-generative AI than we had seen before. It's not that AI is a new problem; t it's just got renewed interest.

[19] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 March 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. A copy of the Act can be found here: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689.

[20] Some of the legal concerns that arise with generative AI include questions over intellectual property rights, copyright protections, personal data and confidentiality, and contractual terms. *The legal implications of Generative AI*, DELOITTE, https://www2.deloitte.com/us/en/pages/consulting/articles/generative-ai-legal-issues.html.

**Elizabeth Townsend Gard:** Interesting, interesting. And Chad, other thoughts about this moment? Where are we? Can you help us understand the historic context? How did we get to where we are, and where are we at the moment?

**Chad Rutkowski**: So I bracket AI into essentially two different buckets: 1) what I call boring AI, and 2) generative AI, which is exciting and all the rage. Boring AI though is what Theresa was talking about, which is the machine learning data driven trying to tease out insights out of your data. Data analytics sort of falls into one of three columns:

- descriptive analytics: You've got a bunch of information.. What is it saying? What is the state of play?
- predictive analytics: We know what the data says. We have some, degree of pattern detection, and we can say that you know the current state is this, but the future state will likely be this, based on, historical information or other factors.
- prescriptive analytics, the gold standard.

I had dealt with that for years in healthcare and detecting population health management trends. There's a spike in diabetes cases in this particular community. Let's deploy nutritional information, and there to try to drive down the hospitalizations or drive down the healthcare costs by proactively giving preventive type treatments. Or Fintech. You look at what the risk profiles are in issuing loans. But at the same time, GPT2 was being created and GPT 3 was being created, at the Copyright Society, we put on a panel.[21] I guess it was in 2020, where we had a Professor Hector Postigo at Temple University, demonstrate GPT.[22] And we created a Yeats poems because I was an Irish literature major. So I asked him, "Can you make some AI generated Yeats poems?" and they were pretty good. They sounded sort of in the style of Yeats.

**Elizabeth Townsend Gard:** Just curious. How did the Copyright Society react to all of that? Because that, I imagine that would be slightly frightening.

**Chad Rutkowski:** Yeah, and there was a fair amount of consternation. And we were starting an AI steering committee at the time and some other folks. Monica Corton and Bhamati Viswanathan really wanted, from the rights holder

[21] On November 23, 2020, the Copyright Society hosted Nadia Baksh, Hector Postigo, Paul Reinitz, and Chad Rutkowski for a panel discussion on emerging machine learning programs and their effects on the legal field. The panel explored how machine learning can transform existing content into new forms and the data markets behind such tools, as well as an examination of relevant legal issues such as fair use and authorship. *Content as Data and Data as Content: Copyright Issues in Machine* Learning, THE COPYRIGHT SOCIETY, https://copyrightsociety.org/event/content-as-data-and-data-as-content-copyright-issues-in-machine-learning/.

[22] Hector Postigo is an Associate Professor at Temple University where he researches new digital media, cultural production, and technologically mediated activism. He is the author of The Digital Rights Movement and the co-editor of Managing Privacy through Accountability. *Id.*

side, the content owner side, to monitor. We had Paul Reinitz from Getty, who was eventually was involved in the *Getty v. Stability* AI lawsuit. He spoke alongside of us with Hector and and Nadia Steele. I had a slide that said, "Data is AI fuel."[23] And that in the AI world, content is data and data is content. And what I meant by that was is that you may not think of stories, or paintings as data. But to a data scientist, they assign different values. Break it down by the pixel, you know. Break it down by paragraphs. The notion of tokens is to try to detect patterns of how language works, how expression works. So to a data scientist, the content is data in that sense. And things like collections of information, for example: temperature data that might be thrown off of an industrial piece of refining equipment or a chiller. It's not content, but when you can collect that together all of a sudden, that collection becomes incredibly valuable and important. So the data sort of becomes content. And when you ask a data scientist, "How much data do you need in order to do their jobs?" The answer is always "more." And we see that with the notion of LLM collapse and data drift,[24] and we can talk in a moment about. But in terms of how did we get here? I would say that you know the innovations that drove things that we've kind of become used to right natural language processing that gives us, things like Siri, and machine learning algorithms that give us autonomous vehicles. There's been an enormous amount of work in this space going on in the background for the past 15 to 20 years. There was a paper that Google had put out, *Attention Is All You Need*, where they talked about transformer models. [25] That sort of set woke the world up to the possibility of transformer models and led to a lot of the academic work that drove open AI at the time, including the GPT2 and 3 models,[26] and the LAION database

[23] The idea that data is the fuel for AI comes from the fact that data is the foundation given to AI engines that enables it to learn patterns, make predictions, and improve over time. AI engines are trained on large datasets to recognize patterns which allow them to adapt to become more accurate and reliable. *Why Data is the Fuel for AI*, COLLECTIVE INTELLIGENCE (Nov. 21, 2024), https://www.collectiveintelligence.com/why-data-is-the-fuel-for-ai/.

[24] The idea of LLM collapse and data drift comes from when AI models are trained off their own output or other generated data, which leads to long-term learning problems. The models then degrade in quality and generative data that drifts further away from fact. This false data can then be used to train other models which will increase the issue and can eventually cause the models to collapse. *New research warns of potential 'collapse' of machine learning* models, UNIVERSITY OF OXFORD (July 25, 2024), https://www.cs.ox.ac.uk/news/2356-full.html.

[25] Transformer models are a type of deep learning architecture that are highly intelligence at processing sequential data by understanding the relationships among the different parts of the input. Ashish Vaswani et al., Attention Is All You Need (June 12, 2017) (unpublished manuscript), https://arxiv.org/abs/1706.03762.

[26] GPT-2 and GPT-3 are generative pre-trained transformer models that were developed by OpenAI in 2019 and 2020. GPT-2 was the predecessor to GPT-3, with GPT-3 being the larger and more capable model of the two. *GPT-2 (GPT2) vs GPT-3 (GPT3): The*

models.[27] And at some level, this was a technology looking for a problem. We know these models can create output. But what is it good for? What are we going to use it for? And so you know the advent of ChatGPT, in November of 2022, that set off a lot of light bulbs that "Wow, this stuff actually works." It's actually: still clunky, but kind of good. And we've just watched this, sort of evolutionary telescoping scaling, this exponential growth has been amazing.. Like Moore's Law, that computing power doubles every year or so. You just have this compression of advancement. That same principle, I think, has occurred in the past two years where the stuff that was kind of good has become really good. And so where I think we're at right now is we're in several places at once. We're in a hype cycle.

I don't think that AI is akin to NFTs. I don't think that AI is akin to Pets.com and eToys.[28] And just like the Internet bubble, I think there's going to be somewhat of a contraction or rolling back, but I don't think it's going to be as severe as what we saw with the Internet in 1999 and 2000, with that bubble bursting.[29] I also don't think it's going to be as severe as what we've seen with NFTs or the "metaverse".

**Elizabeth Townsend Gard:** I'm just going to pause. I'm thinking of people 30 years from now. Can you help us understand what you mean by the Internet bubble and the NFT bubble, and why? It's different. Why, you think this moment is different?

**Chad Rutkowski**: Sure. So in the late 1990s, it was very similar to what we're seeing with AI right now. In the beginnings of the Internet, which was largely a chat room type system set up between university students to be able exchange research papers and then rolled out commercially. And so this means of using computers to have more immediate interaction and then began to find commercial applications. So we had a sort of birth of e-commerce in the late nineties, and that led to enormous amounts of speculation, and enormous amounts of investment that wasn't necessarily borne out in profits. So you had a lot of darlings of the early Internet era, including pets.com, and Yahoo. And a lot of them just collapsed. And it led a lot of people to say, "Oh, the Internet was fake."

---

*OpenAI Showdown*, EXXACT (Feb. 3, 2021), https://www.exxactcorp.com/blog/Deep-Learning/gpt2-vs-gpt3-the-openai-showdown.

[27] The Large-scale Artificial Intelligence Open Network (LAION) is a non-profit organization that provides tools, datasets, and models for machine learning research. LAION, https://laion.ai (last visited May 18, 2025).

[28] *See* pets.com and eToys.com.

[29] In the late 1990s there was a period of rapid growth in internet-based companies due to an increase in technological advancements and investor enthusiasm. This enthusiasm led to a stock surge but many of the surging companies lacked solid business models. The dot-com bubble burst in early 2000 which led to massive fancial losses and the collapse of many startups. *The Late 1990s Dot-Com Bubble Implodes in 2000*, GOLDMAN SACHS, https://www.goldmansachs.com/our-firm/history/moments/2000-dot-com-bubble.

The promise of it all was not it was always cracked up to be. I don't know about you, Elizabeth, but I continue to use the Internet every day…

**Elizabeth Townsend Gard:** We're doing it right–

**Chad Rutkowski:** Right now.

**Elizabeth Townsend Gard:** That's right.

**Chad Rutkowski:** Yeah.

**Elizabeth Townsend Gard:** And that the influx of capital and energy and excitement over it was kind of out of control, it seems. And so you're saying that a lot of those companies didn't make it. And then it kind of settled down. And you're saying the same thing happened with NFTs. I'm curious, when I look at the numbers of what's being invested in AI companies, it is staggering. It's not even conceivable how many billions of dollars a particular company has. And I'm sure that's because of the data and all kinds of other things. But how do we understand the investment that's happening now with AI companies as opposed to those startup Internet companies, that Internet phase? Is it exactly the same? Or is it a little bit different? It seems like there's less of them, but that the money they're getting is just incredible. And they're sprouting up very quickly. Am I right about that? It just seems kind of different.

**Chad Rutkowski:** What I would say is it takes a lot of money to run this stuff, and we'll see how long that stays true. Now that we have new entrants like China's DeepSeek,[30] and others that are challenging that. But the amount of infrastructure computing power that it takes to create AI outputs requires: a lot of the money. And so it does sort of naturally creates a limited number of players, because there's a natural barrier to entry.

**Elizabeth Townsend Gard:** I'm curious. Does it take a lot of money to run it or to build it? I'm thinking of the story of AT&T. It takes them 100 years to get to where they are, and then they get broken up. But Skype can build upon the AT&T infrastructure and thrive in years.[31] Are the companies we see now having to build a lot, and that's where the money is? Or is it because the running of it is so expensive, is it going to get less expensive as we go? How do we understand? Kind of infrastructure versus, like the day-to-day operations of this kind of stuff?

---

[30] DeepSeek is a Chinese artificial intelligence company founded in 2023 by Liang Wenfeng. The company has developed several open-source AI models that have been praised for their efficiency and performance. Kelly Ng et al., *DeepSeek: The Chinese AI app that has the world talking*, BBC (Feb 4, 2025), https://www.bbc.com/news/articles/c5yv5976z9po#.

[31] Skype was a telecommunications application released in the early 2000s that allowed users to communicate with other via the internet. *About Skype*, MICROSOFT, https://www.skype.com/en/about/. As of May 5, 2025, Skype has officially been retired by Microsoft. *Skype is retiring in May 2025: What you need to know*, Microsoft, https://support.microsoft.com/en-us/skype/skype-is-retiring-in-may-2025-what-you-need-to-know-2a7d2501-427f-485e-8be0-2068a9f90472#:~:text=As%20of%20May%205th%2C%202025,same%20core%20features%20and%20more.

**Theresa Weisenberger:** It will get less expensive both to use and develop AI technology as we go on. What Chad said earlier about needing all of the data to create these LLMs, if companies like Deep Seek who are focuses on develop more efficient technology without compromising its functionality are successful,, then we can see a decrease in the cost of maintaining these or running these systems. A big part of it is how much electricity and computing power is needed to operate. It is much more environmentally costly to have prompt generative AI tool to provide search results than it is to run a typical search on a search engine. If the semiconductor chips can become more efficient, and if we can make the LLMs themselves more efficient, I think we will see a decrease in the environmental impact and the cost. But right now we're still in t the early Internet days–right before the.com bubble burst–where we're paying for generative AI tools based on "tokens" used, like when AOL would give you Internet access for a certain number of minutes a month.

**Elizabeth Townsend Gard:** So am I right that 2022 is an important marker, but that things came before that. How do we understand the November 2022 release of ChatGPT?

**Theresa Weisenberger:** That was when AI became tangible to the general public. It was being used before, and of course it was being developed long before that, but AI was something that we really didn't have good language or examples to talk about in ordinary conversation before ChatGPT.

**Chad Rutkowski:** And it marked the changeover from boring AI to fun AI! Right? To generative AI. I can't tell you how many of our clients at the time in late 2022 and early 2023 told us, "Oh, we want to use ChatGPT to sort resumes or detect cancer. Help us manage and use ChatGPT." Now, maybe now, right there, there's different services and plugins. But they didn't realize that "You're talking about the boring AI that you've been using for the past ten years." So what we dealt with a lot in early 2023 was this conflation of AI. We spent a lot of time having conversations like this. 'Let's slow down. Let's talk about what you're really doing, and that you're not even using ChatGPT, that's not what you mean, even though that's what made you think of it." That moment at the end of 2022, and the first quarter of 2023, we grabbed a lot of people's imaginations, made AI interesting and required a lot of education.

**Elizabeth Townsend Gard:** Okay, then, 2023, we get the lawsuits. It's like the year of the lawsuits. Help us understand what's happening. Did the authors and the artists, and everyone not know? Were they just becoming aware of Generative AI? Is that why it happens in 2023? Why in 2023 do we see this huge amount of lawsuit litigation happening?

**Chad Rutkowski**: Oh, I know exactly– money. Because you know, we saw investment. Microsoft investing in OpenAI. We've seen Google and Amazon investing in Anthropic. And some of these other companies. We've seen Google trying to create its own versions, Meta trying to create its own versions. So suddenly, there was commerciality to this, and potential deep pockets. And

whereas beforehand it was all this is nothing commercial, a lot of research communities and nonprofits like Common Crawl.

We know what the vast bulk of data that GPT and certainly GPT 2 GPT 3 were trained on. There are a lot of claims that the AI companies are so non-transparent about what they trained on. But in 2019 US Patent and Trademark Office issued a request for comments on all things AI and IP.[32] Initially it was focused on patents, and that was because Steven Thaler claimed in the summer of 2019 that AI had created a patent for him, and it set the world ablaze right like that was the first AI hype cycle, even before ChatGPT came out. His assertion that AI should be recognized essentially with human-like rights and a patent should issue as if AI was an employee, an agent, as the basis of ownership.[33] So in that request for comments, OpenAI responded to that and explained what they were doing and laid out the basis for most of the data set training data sets in Gpt3, which included Common Crawl.[34] Common crawl had lots of books in it, and there was no mystery to that. That's again at the time that I was giving these presentations with the Copyright Society,[35] but I also gave them to the American Bar Association. Both the Patent and Trademark Office and the Copyright Office held AI days as well.[36] They had day long presentations about these issues. At the time, the Rembrandt project was at the height of AI technology.[37] That they could create new Rembrandts because the AI was trained on Rembrandt paintings. And what was this going to mean for new tools coming forward. There was a lot of interest in it. A lot of it was sort of academic interest. A lot of it was trying to understand what the industrial business use cases were going to be. And you know there were a number of people on top of it from the get-go. One of the things that

[32] Request for Comments on Patenting Artificial Intelligence Inventions, 84 FR 44889 (Aug. 27, 2019), https://www.federalregister.gov/documents/2019/08/27/2019-18443/request-for-comments-on-patenting-artificial-intelligence-inventions

[33] Thaler v. Vidal, 43 F.4th 1207 (Fed. Cir. 2022), https://www.cafc.uscourts.gov/opinions-orders/21-2347.OPINION.8-5-2022_1988142.pdf.

[34] Common Crawl is an engine that crawls the web and provides its archives and datasets to the public for free. Mission, COMMON CRAWL, https://commoncrawl.org/mission. A copy of OpenAI's comments to the PTO can be found here: https://cdn.openai.com/policy-submissions/OpenAI%20Comments%20on%20Intellectual%20Property%20Protection%20for%20Artificial%20Intelligence%20Innovation.pdf.

[35] *See Content as Data and Data as Content*, *supra* note 24.

[36] The Copyright Office and the World Intellectual Property Organization co-sponsored a symposium on Copyright in the Age of Artificial Intelligence in February of 2020. Information and recordings of the event can be found here: https://www.copyright.gov/events/artificial-intelligence/.

[37] The Rembrandt Research Project is a collaboration by a group of art historians to produce a comprehensive catalogue of Rembrandt's creative works as well as research on the artist himself. *About Us*, REMBRANDT RESEARCH PROJECT, https://www.rembrandtresearchproject.org/about-us/.

I did for the American Bar Association at this time is help the ABA write its response to the PTO's request for comments in 2019,[38] very similar to what the Copyright Office did in 2023, with, its notice of inquiry. And Mary Rasenberger of the Authors Guild, Josh Simmons from Kirkland, Steven Chow, and Dave Green, who is now at Github, worked on it with me, and we wrote a position paper to the PTO.[39] So we knew this was coming. We knew that adoption of these tools was going to occur rapidly. It was the roll out of a commercial product that starts the lawsuits.

**Elizabeth Townsend Gard:** Yeah, that makes, I mean as an outsider. That also sort of makes sense. I mean, if you're bragging about the money that's coming in. People are going to start to feel like, "Hey, what? You're using my stuff and that's not fair. And why don't we get licenses?" So 2022 is the public recognizes Generative AI, and by 2023, many lawsuits are filed.[40] Theresa, when does your firm start tracking these Generative AI lawsuits, and help us understand where the tracker comes in? You've got one of the best trackers out there for AI cases.[41] Related to this, how many cases are there? We can't figure out how you count them. We've decided that there's no way to actually know how many they are, because everybody counts them differently. Is that true? And tell us a little bit about that.

**Theresa Weisenberger:** That is definitely true. I looked at the beginning of your working draft of your article [AI Litigation v. Licensing], and it shows how many different ways to count how many cases there are. I can't figure it out either, and a lot of it is more art than science.

**Elizabeth Townsend Gard:** That's so amazing. Usually, you're like, "Well, how many cases are there?" 42. But in this case, it isn't because they keep getting consolidated, but everybody has their different numbers and which ones they care about, and which ones get ignored. It's really weird.

**Theresa Weisenberger:** Not only are they getting consolidated, but we've also seen copycat cases filed. In *Tremblay v. OpenAI* and *Silverman v. OpenAI* .The complaints are almost identical. The plaintiffs are different, and so the value of tracking those cases was a judgment call as far as bandwidth and everything else. And eventually those cases ended up getting consolidated into In re: *OpenAI ChatGPT Litigation*. And then it was a little bit easier to track.

I started this case tracker in 2023. I'm looking at the website now, I think when I started it, there were 3 or 4 cases filed. I was getting asked by my clients,

---

[38] A copy of the ABA's comments can be found here: https://www.aallnet.org/wp-content/uploads/2020/01/RE-IP-Protection-for-AI-Innovation-010720.pdf.

[39] *ABA responds to Copyright Office Notice of Inquiry on AI and Copyright*, BAKERHOSTETLER (Nov. 6, 2023), https://www.bakerlaw.com/insights/aba-responds-to-copyright-office-notice-of-inquiry-on-ai-and-copyright/.

[40] *See generally* Drew Berg, *Exploring Generative AI Lawsuits Timeline*, TRIALLINE (Mar. 14, 2024), https://blog.trialline.net/generative-ai-lawsuits-timeline/.

[41] *Case Tracker*, *supra* note 5.

"What is this case? What are they focused on? What data sets are they focused on? And what are the theories of the parties, and how are the courts evaluating them?" They weren't all straightforward. We've seen a lot of focus on CMI claims under Section 1202(b). We've seen some interesting common law claims that have sometimes survived motions to dismiss and sometimes been found to be preempted by the Copyright Act. When I started the case tracker, it was because I kept going back to PACER over and over, looking at these complaints, looking at the motions to dismiss, and I wanted a central place to keep them. And I thought it would be good for clients to have access and the public to have access to. So that's it. It really is designed to be a neutral reporting.

**Elizabeth Townsend Gard:** It is, and it's amazing. When you look at our paper, we just keep quoting your site like we just like I had to keep telling them you have to go see other stuff like you can't just do this one site because it's you. Do you feel that it's trustworthy? It's thorough. It's up to date. It understandable. Having now engaged in it with the students for a couple of months, it's been really interesting to see.

**Theresa Weisenberger:** That was intentional. I've got a blog that covers all of these different cases, and as more cases were filed, we needed to include them without running afoul of any existing relationships with the parties involved. So we intentionally summarize, but do not opine on, the case developments. When my clients were asking the implications of these cases to their businesses. I wanted them to come directly to me.

**Elizabeth Townsend Gard:** I hope you continue it. And I hope we preserve it somehow. Because I think it's a really important moment of tracking what's happening. So all of this litigation that begins in 2023. I imagine we've never seen anything like this in copyright. Am I right that, like the level of like the flurry of copyright infringement cases, and the CMI stuff which has never really been a thing in any significant way. This is unprecedented in copyright. Have we seen this before?

**Chad Rutkowski**: The Perfect 10,[42] the thumbnail cases and some of the early Internet cases…

**Theresa Weisenberger:** And the class action aspect is unique. There have been copyright class action cases before this, but not proliferation that we've seen with AI.

**Elizabeth Townsend Gard:** Right. Like all the copyright professors, we have to know about class actions and all the class action professors are like we have to know about copyright. We're not happy about this. You all stay in your lanes, you know. It makes our jobs harder. Yeah, it's pretty funny Okay. So 2023 lots of lawsuits, 2024 licensing deals: the amount of licensing deals that are happening was staggering. What do you think of this turn? So I was talking to copyright lawyer Eric Schwartz yesterday about historical stuff on the copyright law, and he said some of the things that got put into the copyright law were

[42] Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146 (9th Cir. 2007).

bargaining chips for negotiations. Do you think the lawsuits were bargaining chips for licensing deals and how do we understand what occurred in 2024?

**Theresa Weisenberger:** When these cases first started getting filed in early 2023, we did work for clients interested in procuring content licenses from publishers. Licensing content for LLMs raises a whole host of issues that neither the publishers or the LLM providers were in a position to answer. The potential licensors wanted use restrictions, as they didn't want the model to end up being used to create downstream competing products. But that's not how LLMs work, so potential licensees were not able to agree these restrictions. We did see a stall in communications. I do think that for some of the big plaintiffs, ultimately licensing their content was probably the goal. They knew that this technology wasn't going away, so they can at least make sure that they got some type of compensation. For some of the other plaintiffs, I don't think that's true.

**Elizabeth Townsend Gard:** Yeah, I mean, I think what's really interesting is you get these sort of the actor/writer/artist side of it "You're going to replace me." You got all this all the SAG/AFTRA/Writer's Guild strike stuff, which we haven't even talked about. What struck me was when Openai was funding four newsrooms for Axios. AI money becomes the Savior potentially of news organizations. That was that I was not expecting that part that, like that, was a kind of flip in the story that seemed kind of interesting to me. Am I getting the story right? Is that kind of the narrative now of like these licensing deals are going to support ways of allowing news to continue?

**Chad Rutkowski:** It, I think so. Let me also hasten to add that I'm not sure whether even to call these things licenses is a pain point and a bargained for term. Because on some level, to call it a license is to acknowledge that you need a license as opposed to an access right. And there's plenty of sort of academic articles out there that suggests that fair use is not a defense. Rather, it's an affirmative right.

**Elizabeth Townsend Gard:** That's right, and that is all still working its way through the courts. So we have this notion. It's this, I was talking to my students. It feels like the litigation is kind of moving the sway [slowly], and then what the licenses are just whipping through. The newest case kind of really got me, which was Advanced Local Media v. Cohere.[43] In that case the publishers are saying, "Look, we already have licensing deals. Now you're infringing because we have the licensing deals." It was like this really interesting moment of like, making the licensing deals and then using it as a litigation strategy to sue people that didn't get a license for you from you.

**Chad Rutkowski:** And that's why there was resistance. There's a real issue here. As to you know who the proponent of the access. AI is always going to be hungry for more and more content to protect against what we call LLM collapse or data drift, where we get less accurate answers. Where the reasoning is impeded, there's always going to be this sort of need for dynamic ingestion. So I don't know.

[43] Advanced Local Media v. Cohere, No. 1:25-cv-01305 (S.D.N.Y. filed Feb. 13, 2025).

That realistically, from a technological perspective, that scraping and crawling and access to data in the wild is ever going to stop. I don't think that these business deals are going to sort of solve the AI problem. I think they're going to be part of the ecosystem, but they're not going to be the only part. And so you have to look at what the sort of business case the business logic, the business proposition is for controllers of content and AI developers and AI companies. And I think there's a couple of those steel points. You know. One is avoiding litigation. Another is, having access, the more that controllers of content: can provide data that is clean, that is usable, that is free from bias that is robust. That actually is a value. And for AI companies that means something that's worth paying for. And there's public opinion That, as you're saying, Elizabeth, AI as the Savior of the news organizations that that's going to play into it as well.

But I think one thing that we all need to be mindful of is there's a difference between a copyright owner and the controller of content. I'm using that term purposefully right. The controller of content might be an aggregator like a Getty images. That, like a music publisher might be a platform right. The Copyright society in 2 weeks is going to have one session on these content deals in the AI space, and one that we're hoping we believe one of the speakers is going to be Reddit. And Reddit's interesting. They don't own anything. They don't purport to own anything that their users create, but they're licensing access right to that content. So you know it, it matters that's going to matter in the fair use context, Like, for these class action suits. That is a collection of individual artists, and the value prop for licensing one author's work is enormously different from getting access to a data set that might be owned by Getty. Right. Getty may not have rights to the underlying photographs, but wouldn't necessarily own them. So are you a compilation owner? Are you an individual content owner? You know an individual work owner rather. Take JK Rawling where her books are often held up as prominent in some of these AI models. If AI models spit out a Rawlings book, well that's bad data science. It's called memorization. That's just because it wasn't deduped.[44] It doesn't actually have any real value. No, AI company is going to say, "Oh, you're JK Rowlings. Let me give you 10 million dollars because you're going to teach my AI model so well that you know it's worth more." It's not worth different than any other collection of 800 or 8,000 pages of literature.

**Theresa Weisenberger:** The Harry Potter books are some of the most frequently pirated content on the Internet. So that's one of the reasons why it's hung up as an example over and over is because multiple copies of it have been included in the data sets that are used to train these models.

**Elizabeth Townsend Gard:** Interesting.: Do you think we'll ever see copyright holders, the photographers suing the platforms like Getty that they're making these big deals?

[44] 'Dedupe' is an abbreviation for 'deduplication,' which is when unnecessary copies or parts of a file are eliminated from a dataset. *Dedupe*, CEGAL, https://www.cegal.com/en/dictionary/dedupe.

**Theresa Weisenberger:** I think that we've seen that in *Anderson v. Stability AI*. DeviantArt was a defendant based on what they were doing with user-uploaded content: One of the arguments was that they had changed their terms of use, so that they had the right to do with your content as they did. And you know there's debate over whether they provided adequate notice and whether the training data went back to content that existed before then.

The most interesting case that has addressed this issue or that will address this issue is *Pierce v. Photobucket* in Colorado.[45] Photobucket was where everybody put all of their images maybe ten years ago.[46] And its really kind of become this undervalued graveyard of photographs. Now that we're looking at online content differently. It is so valuable to AI development. Photobucket is all of a sudden sitting on a lot more valuable asset than they were a few years ago, and they have been sued by a few [of the photographers] as a putative class action. Photobucket was sued, not on the basis that it has actually licensed user content yet for AI development, but that plaintiffs believe based on recent Photobucket activity that such licensing it's coming down the pike.

**Elizabeth Townsend Gard:** Fascinating, fascinating. Okay, I know we've been here already for a while. Let me ask you 2 quick questions. Help us understand what your thoughts are on fair use arguments. I'm curious about the *Ross* case, and I'm also curious about the cases that are doing the 1202 decisions that are coming out, and how you all are thinking about that.

**Theresa Weisenberger:** So on the *Ross* case,[47] it was hugely important that copyright infringement in the context of creating AI models was an issue of first impression for any court. And these summary judgment decisions were particularly significant, because after first denying summary judgment, the court invited the parties to rebrief the issues, and ultimately the court changed its mind. But how the recent summary judgment decision is being interpreted and how it's going to end up affecting these other cases is going to be a lot more interesting. As the opinion states, what we're talking about in *Thompson, Reuters v. Ross* is not generative AI. It's not necessarily a one-to-one comparison about whether one use is fair or not. The court found that they were creating a competing product, unlike t these other cases. There isn't some type of assumption that Stability AI is creating a competing product to the content owners.

**Elizabeth Townsend Gard:** Okay. Let's pivot. The Copyright Office has come out with two out of now three reports relating to AI. I'm curious what your

[45] Pierce v. Photobucket, Inc., No. 1:24-cv-03432 (D. Colo. filed Dec. 11, 2024).

[46] Photobucket is a company founded in 2003 that offers online backup cloud storage services for digital media. *Photobucket Reviews, Ratings & Features 2025*, GARTNER, https://www.gartner.com/reviews/market/photo-management-software/vendor/photobucket/product/photobucket#.

[47] Thomson Reuters v. Ross Intel., No. 1:20-cv-613-SB, 2025 U.S. Dist. LEXIS 24296 (D. Del. Feb. 11, 2025).

thoughts are. Did we learn anything new? There's a lot of effort being put into this. So I'm curious what you think.

**Chad Rutkowski:** When the first report came out, it almost seemed coordinated, although I think we've been assured that it wasn't, with the legislation that that was positive at the same time. And I think it really spoke to concerns over deep fakes and the need to address digital replicas.[48] But in this political environment, I just don't see you know any meaningful legislation going forward. So it's to me the first report was more a policy statement or a clarifying statement that this needs to be handled by Congress and you ought to get on this.

**Elizabeth Townsend Gard:** Do you think that because of our political climate, at the moment that the States will take it up that we'll see more state engagement in the deep fake problem.

**Theresa Weisenberger:** Yes. I was looking at the deep fake problem, which has been on states' radars at least as far back as 2021. Louisiana that had laws specifically targeting deep fakes in certain criminal contexts.[49] But with this surge of AI we've seen j that the federal government is not in a position right now to actually meaningfully legislate, at least on these topics. We have seen step states step up to address AI issues the same way they stepped up to address data privacy, like California did with the California Consumer Privacy Act.[50] We will continue to see states tackle AI issues in similar ways. Last year we saw a ton of bills proposed, and, I want to say, maybe a third of the states didn't pass [proposed AI legislation],[51] and we're as early in the legislative season for 2025, but new AI bills have already been proposed. A lot of the states are also going to follow on to what we've seen happening with the EU AI Act.[52] Parties are looking towards the language and the restrictions of the EU AI Act not only to comply with the EU and future state AI legislation,, but also incorporate similar restrictions and rights into their contracts in which they are either providing or using AI

[48] The term 'digital replica' can refer to an image, video or audio recording that has been digitally created or edited to depict an individual, also known by the term 'deepfake.' Digitial replicas can be both authorized or unauthorized content. Alexander Touma et al., *Overview of the U.S. copyright office's AI digital replicas ("deepfakes") report*, A&O SHEARMAN (Aug. 2, 2024), https://www.aoshearman.com/en/insights/ao-shearman-on-tech/overview-of-the-us-copyright-offices-ai-digital-replicas#. The first report on AI from the Copyright Office can be found at the following: U.S. Register of Copyrights, Copyright and Artificial Intelligence Part 1: Digital Replicas, (July 2024), https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-1-Digital-Replicas-Report.pdf.

[49] La. Rev. Stat. § 14:73.13 (2023), https://www.legis.la.gov/legis/Law.aspx?d=1336015.

[50] Cal. Civ. Code § 1798.100 et seq.

[51] *See Tracking U.S. state artificial intelligence* legislation, HUSCH BLACKWELL, https://www.huschblackwell.com/2024-ai-state-law-tracker.

[52] *See* Regulation (EU) 2024/1689, *supra* note 22.

technology, so even if the states don't legislate, the parties can agree by contract how the AI issues should play out.[53]

**Elizabeth Townsend Gard:** Interesting. I love it. And then part two from the Copyright Office?.

**Chad Rutkowski:** I have a lot to say on that, and you know I'm working on a blog article to address it. I

I'm looking forward to the Copyright Office's revised guidance on registration. You know that was hinted at in the second report. I was a little surprised to there was quite a lot of emphasis and reliance on Professor Jane Ginsburg's law review article, "Machines and Authors"[54] and she had a co-author. A good bit of that article focused on a type of copyright authorship. An author is not just this sort of artist in the garret that has this brilliant idea spring from the top of their head and end up on a page–that's classic authorship and certainly one form of the author. But the law recognizes other types of authors. Obviously, compilation authorship. But what I find interesting is what we call mastermind authorship and it's why we give authorship rights, ownership rights to production companies right in the *Al-Mohammed* case, the *Casa Duce* case. These are entities that may not have had a single creative thought, but they superintended the arrangements. They brought together the various creative elements. As a hallmark of some of those cases is what I call a veto. There's a major focus in the *Kashtanova* registration decision.[55] And again, in this second report over control over predictability, you'd never know what the AI is going to hand back to you. At one point, it was described in the report as spinning a roulette wheel over and over again, and: there may be some truth to that. But it's kind of true with humans, too. They never know what they're going to hand back to you. And you know the production studio has the veto. This goes in, and that doesn't right. I get to decide what the final product looks like, whether to accept it or not, whether fire the director, or to bring another scriptwriter.. The actors–bring another set of actors on until I get it right, and it meets some degree of mental conception. The production company wants the film to be what it wants it to look like. And you know I, personally don't see any major difference in that type of iteration with what we have in a Theatre de Opera. You know where they iterated 600 times in MidJourney to get to that concept. And I was disappointed that you know there's

---

[53] *See generally Model Contractual Clauses for AI Procurement in the EU: Key Takeaways for AI Companies*, Cooley (Mar. 20, 2025), https://cdp.cooley.com/model-contractual-clauses-for-ai-procurement-in-the-eu-key-takeaways-for-ai-companies/.

[54] Jane C. Ginsburg & Luke A. Budiardjo, *Authors and Machines*, 34 BERKELEY TECH. L. J. 343 (2019).

[55] The US Copyright Office granted Kris Kashtanova's graphic novel "Zarya of the Dawn" limited copyright protections, which was a signficiant step in the development of copyright standards for works created through generative AI platforms. Tom Hals & Blake Brittain, *Insight: Humans vs. machines: the fight to copyright AI art*, REUTERS (Apr. 1, 2023), https://www.reuters.com/default/humans-vs-machines-fight-copyright-ai-art-2023-04-01/.

a lot of discussion about joint authorship. And how the cases are analogous to the situation. But no discussion about mastermind authorship, even though that was in the Ginsburg law review article. The other thing that I found a little disappointing is that it seems like Mid Journey is driving all the policy here, that image generation, and you know the concerns of visual artists which are super important, but not the only use. You look at. from what I understand, for what Theresa and I think, see in our business, and from what we benchmark with other practitioners, 80% of the adoption. The real uses of these tools are in code generation, in source code, software code generation. So to have no discussion and the authorship rights in this context [is disappointing]. When you're using CodeWhisperer, you're using Copilot is problematic, because what I think it does is set up for a situation of civil disobedience. If I disclaim the AI authorship here, that's you're done right, you're not able to unring that bell or I can decide to take on the expense of fighting the copyright officer. The PTO. I think, takes that approach right. The PTO seems like it has a much more deferential approach, and you'll fight it out in court. And we deal with this all the time. When I give these talks I take my phone out and I'll go like that. And I'll say, do I have a copyright in the photograph?. You know we're going back to *Burrow-Giles*.[56] We're going right back to that. I didn't pick the lighting. I didn't position you. I own the phone. Am I the author? I don't know but you know what I'm going to get: I'm gonna get a registration on that photograph. I want to spend the money to enforce it. It will get this determined by the courts as to whether or not I'm an author.

**Elizabeth Townsend Gard:** I don't know. It just starts to really unravel part of the physics of copyright that we all know so well by not really addressing it at the moment. So I find it interesting, and I like the idea of the mastermind that we are.

I have just loved our time together. You are both the two of the smartest people on the planet on this stuff, and we're so lucky that the Copyright Society has you as the members and so involved, and that you're willing to have an interview, and that we can do this. So anything else before we end that we haven't covered that you'd like to cover, because I mean again, we could talk forever.

**Chad Rutkowski:** I'll just as quickly as I can, you mentioned *Thomson Reuters v. Ross*.[57] I have a blog article out on it, but you know one thing that threw me for a loop a little bit was on the derivative argument.. The infringement that supposedly occurred was there was a creation of bulk memos where the head notes were reproduced entirely, and then a database created from that in order to train…

**Elizabeth Townsend Gard:** And to compete with the originals.

**Chad Rutkowski:** I think that you know this is going to sound convenient, but I think a positive: fair use decision would have been very problematic for content owners. But I don't think that this decision is equally problematic for AI

[56] Burrow-Giles Lithographic Co. v. Sarony, 111 U.S. 53 (1884).
[57] *See* Thomson Reuters v. Ross Intel., No. 1:20-cv-613-SB, 2025 U.S. Dist. LEXIS 24296 (D. Del. Feb. 11, 2025).

developers. And the reason is because the general purpose nature of large language models. This is what I call a small model. Competition here was direct. One of the things I think some of the AI cases suffer from is this notion you're creating a competing product and materials. Well, at what level? It's not that competition problem is with specific works. If you create an AI model, you know that creates the specific same work and competes. That's the thing: if it's competition in the vague, sort of universal sense of it, well, that's not what the case law focuses on.

You know one thing that I think was missed in the decision, and I think it has got to become more prominent is this notion of extraction. You know we have a case wire data from 2005 about taking a database and downloading it temporarily to extract unprotectable information. I think that there's a corollary between what data scientists do with content, although, I don't mean it to minimize the importance of creativity, and I think would appall most people to think of their works of art as just simply being containers of information. There is a fair use argument to be made that if you're simply downloading the express content temporarily to extract unprotectable patterns, facts, information that could be fair use. There's a really interesting journal article called "The Freedom to Extract and Copyright Law," written by Molly Schaeffer von Howling just this past month.[58] That sort of collects those cases, addresses it. So that's you know that that to me is one of the things I think is most interesting about *Ross*.

**Elizabeth Townsend Gard***:* So interesting, and there really hasn't been any kind of conversation, I think, about the *Salinger* case of the unauthorized derivative novel, Sixty Years Later. It'll be interesting to see what happens the moment that someone tries to create their own JK. Rowling novel and sells it, using generative AI tools.

**Chad Rutkowski:** I hasten to add to and to amplify the point that you're making, Elizabeth, which is where you look at the technology matters. So we know about large language models, foundation models. Think of those as like sort of, you know, the giant flywheel. And then you have more fine tuning, you have rag. The better use analysis in a general-purpose foundation model is likely to be quite different than when you are fine tuning, or you're using rag or tag, or any one of these different types of training models. So you really need to understand the technology, you need to understand what's happening. You know what the alleged infringer did, how they did it. I think that's you know it's going to become very grand.

**Elizabeth Townsend Gard:** But so important.. So our family made a fake Christmas letter and then asked ChatGPT to put it in the style of Sally Bowles and it knew exactly who Sally Bowles was, and we didn't have to say from *Cabaret.* And again, it sort of makes me think about the protectability of characters and

[58] *See* Molly S. Van Houweling, *The Freedom to Extract in Copyright Law*, 103 N.C. L. REV. 445 (2025), https://scholarship.law.unc.edu/cgi/viewcontent.cgi?article=7009&context=nclr.

derivative works and all that. And what you're saying, what does these systems have to do on the back end to not have that a thing. And, as you're saying, what is the responsibility of the technology in that Sally Bowles Christmas letter? And what is on the person that asked for it?

**Theresa Weisenberger:** And how much of that style is protectable? Copyright law does not prohibit me learning to imitate an artist's style–

**Elizabeth Townsend Gard:** But what is the style? When is it style, and when is it expression? I don't know the answer to that right.

**Theresa Weisenberger:** If we could correctly provide the answer to that, I we would be very wealthy lawyers…

**Elizabeth Townsend Gard:** Right, totally.

**Theresa Weisenberger:** And I think that we have seen similar issues addressed in the cases evaluating the public domain status of Sherlock Holmes..[59] We've got Sherlock's Holmes as a character in the public domain at one point, while some of the later-developed traits of his character remained subject to copyright. of his characteristics, how much of those actually are in the public domain, and how much of them were created after the fact. When you start slicing up copyright protection of a character whose personality evolved over a long period of time, to determine whether an allegedly infringing work portrays the public domain characteristics of Sherlock or those still subject to copyright, it gets muddled very quickly. .

**Elizabeth Townsend Gard:** Right. But the LLMs aren't distinguishing between something that's in the public domain and Sally Bowles, that is still under copyright..

**Theresa Weisenberger:** Oh, no! I'm just saying what part of a character is protected will be challenging. These are analogous to issues AI developers face when separating out protectible and unprotectable elements of copyrighted works that may be included in AI outputs. Researchers have experimented with how much the AI outputs can be restricted–like refusing to answer prompts that blatantly ask for infringing content (like the text of a news article)--without being overly restrictive, like the filtering of raw facts. And there are harder questions when generative AI is used to analyze a work–while literary criticism is squarely within the context of fair use, for example, it's hard to quantify when discussion of a work rises to the level of infringement.

**Elizabeth Townsend Gard**: Thank you again for this interview. You're a great team, and you're both amazing. And thank you again for all your time. I really appreciate it.

**Chad Rutkowski**: You bet! Thanks for inviting us. Take care!

---

[59] Carter Ostrowski, *Sherlock Holmes: To What Extent Can a Character's Feelings Be Copyrighted?*, UNIVERISTY OF CINCINNATI LAW REVIEW, https://uclawreview.org/2020/11/18/sherlock-holmes-to-what-extent-can-a-characters-feelings-be-copyrighted/#_ftn1.