

**Copyright + Technology Conference
Copyright Society September 30, 2024**

**PANEL 1
MY BLANKET AND ME: BLANKET LICENSING
FOR GENERATIVE AI**

with DAVE DAVIS,¹ RACHEL FERTIG,² UMAIR KAZI,³ ROANIE
LEVY,⁴ AND PHIL SHERRELL⁵

This first panel reflects parallels that the current AI era has with the development of interactive streaming two decades ago: the potential need for online services to take licenses to “all” content of a given type, and thus the need for mechanisms to enable such licensing on a massive scale. Solutions to this problem for streaming developed exclusively in the market until a statutorily mandated solution – for music, in the form of the Music Modernization Act – appeared in 2018. Analogously, this panel discussed various companies’ efforts to build blanket licensing platforms for AI uses in advance of any substantive results from courts or Congress.

Phil Sherrell: I’m Phil Sherrell. I’m a copyright specialist at Bird & Bird in London in the UK.⁶ As I nearly froze on my overly air-conditioned flight from London to New York yesterday, I was reflecting on the role of blankets. Blankets brought me comfort and warmth yesterday. But, of course, they also have another purpose of depriving fires of oxygen and putting out the flames. And it’s possibly

¹ Dave Davis is the co-founder and CEO Calliope Networks. *Dave Davis*, THE COPYRIGHT SOCIETY, <https://copyrightsociety.org/bio/dave-davis/> (last visited Dec. 13, 2024).

² Rachel Fertig is currently a Partner at DLA Piper and serves as the DC Chapter of the Copyright Society. *Rachel Fertig*, DLA PIPER, <https://www.dlapiper.com/en-us/people/f/fertig-rachel> (last visited Dec. 13, 2024).

³ Umair Kazi is the Director of Policy & Advocacy at the Authors Guild. *Umair Kazi*, THE COPYRIGHT SOCIETY, <https://copyrightsociety.org/bio/umair-kazi/> (last visited Dec. 13, 2024).

⁴ Roanie Levy is a Licensing and Legal Advisor for the Copyright Clearance Center. *Roanie Levy*, THE COPYRIGHT SOCIETY, <https://copyrightsociety.org/bio/roanie-levy/> (last visited Dec. 13, 2024).

⁵ Phil Sherrell is head of Bird & Bird’s London office and Partner in their Media, Entertainment and Sports team. *Phil Sherrell*, BIRD & BIRD, <https://www.twobirds.com/en/people/p/phil-sherrell> (last visited Dec. 13, 2024).

⁶ Bird & Bird is a large international law firm focusing on IP and technology. which represents about two-thirds of their business and is really at the heart of the firm. Bird & Bird has over 1,600 lawyers in 32 offices across Europe, the Middle East, and Asia Pacific, but is not currently practicing law within the U.S. For more information regarding Bird & Bird, *see* ABOUT, BIRD & BIRD, <https://www.twobirds.com/en/about> (last visited Dec. 17, 2024).

slightly stretched, but I think we can make a good case that the blankets of a different kind that we're going to be talking about today fulfill all of those functions as well.

Blankets provide comfort to rightsholders in that they know that their content is being licensed on standard terms, and they're getting paid something. Blankets also give rightsholders that warm feeling of knowing that they're getting paid on the same basis as other rightsholders in the same market, and not worrying about their competitors doing direct deals behind the scenes and getting paid more than them.⁷ And I think there's even a case to be made that the emergence of blanket licensing for generative AI might put out the numerous litigation fires that have been started around America between copyright holders and GenAI companies, and also, of course, in the UK and elsewhere.⁸ And we'll touch specifically on that bit of the function of blankets later on.

Dan mentioned earlier about the balance between the interests of copyright holders and the tech companies. I think there's a good case to be made that the UK, of developed markets, is perhaps one of the least balanced in the sense that we have a very, very narrow TDM exception compared to Europe and fair use, and obviously very strong copyright laws. So, good news for rights holders, less so for AI developers perhaps.⁹

Like any good lawyer, I wanted to start this panel by defining a term, and it's that word "blanket" again, just to be clear about what we mean with that. The term is usually associated with collectives, licensing groups of rights for particular use cases.¹⁰ And in the UK and the EU for a long time, we had monopoly collectives that would aggregate all of the rights for a particular copyright work for a particular use case, like public performance of music, and they were the only show in town.¹¹ We're not solely talking about those sorts of traditional collective

⁷ Occasionally, rightsholders may offer lower terms to direct licensees than those in a blanket license to eschew the "cut" that a collective management organization ("CMO") may take, with the result that the rightsholder nets more than it would from the CMO.

⁸ For an ongoing summary, see Baker Hostetler's AI Case Tracker at *Case Tracker: Artificial Intelligence, Copyrights and Class Actions*, BAKERHOSTETLER, <https://www.bakerlaw.com/services/artificial-intelligence-ai/case-tracker-artificial-intelligence-copyrights-and-class-actions/> (last visited Dec. 17, 2024) [hereinafter "Case Tracker"].

⁹ For more information on the legal implications of text data mining ("TDM"), see SCOTT ALTHAUS ET AL., BUILDING LEGAL LITERACIES FOR TEXT DATA MINING (University of California, Berkeley ed.).

¹⁰ These are known generically as collective management organizations (CMOs). Perhaps the best-known of these are societies that collect royalties on public performances of musical works; these are known as performance rights organizations or PROs. Examples of PROs in the U.S. include ASCAP and BMI. See generally *supra* note 7.

¹¹ In July 2008, The European Commission issued rulings to mitigate monopolistic practices of music royalty collecting societies in the European Union, such as to allow songwriters to sign up with the collecting society of their choice rather than restricting them to signing up with the single collecting society in their home countries, and to enable

licensing organizations today, but they are one of the types of licenses we'll touch on.

We're also going to talk about some new market entrants that have formed for the first time, aggregating rights specifically to create new GenAI licenses.¹² And what both of those types of licensors have in common is that they are, of course, representing multiple rights holders and licensing multiple licensees on standard terms. And that's to draw the distinction with the direct deals between individual rights holders and individual GenAI companies, which we're not talking about, not least because they're almost all private and people are therefore unable to talk about them.

So, just in terms of the structure we're going to follow, we're going to shortly introduce a couple of specific GenAI licenses with the help of our panel. Those are sort of illustrative, if you like, but we thought it was helpful in this field, which is obviously very new, to ground the discussion in something specific, some specific licenses for particular use cases. We're then going to move on to a discussion around the complexities involved in formulating blankets in this space. We'll talk particularly about the question of what they cover, the scope of the licenses, and then also a bit about how they're priced. And then we're going to finish off by looking ahead and talking about the catalysts for change, or what's coming up ahead in this market.

So, let's get straight into it. I'm going to ask each panelist to introduce themselves when they first talk. So, let's start by welcoming Roanie from the Copyright Clearance Center. Please introduce yourself, Roanie.

Roanie Levy: I'm Roanie Levy, and I am working with the Copyright Clearance Center (CCC) as an advisor and consultant on all things AI and copyright.¹³ But before that, I was the CEO of Access Copyright, which is a copyright management organization for text-based works out in Canada, and I was there for over 20 years.¹⁴

So, a little bit about CCC. The Copyright Clearance Center has been around for over 45 years.¹⁵ The CCC's mission is to offer solutions that ensure that creators and publishers are protected, while also encouraging and enabling

broadcasters to negotiate pan-EU licenses with single collecting societies instead of having to negotiate with separate collecting societies in each EU member state. *See, e.g.,* David Gow, *EU to introduce new music rights system despite lobby*, THE GUARDIAN (July 14, 2008), <https://www.theguardian.com/technology/2008/jul/15/digitalmusic.digitalmedia>.

¹² *See, e.g.,* Bill Rosenblatt, *The Media Industry's Race To License Content For AI*, FORBES (July 18, 2024), <https://www.forbes.com/sites/billrosenblatt/2024/07/18/the-media-industrys-race-to-license-content-for-ai/>.

¹³ *See* Biography for Roanie Levy, *Understanding AI: More Than Just a Bunch of 1s and 0s*, COPYRIGHT CLEARANCE CENTER: BLOG ON ARTIFICIAL INTELLIGENCE (Nov. 12, 2024), <https://www.copyright.com/blog/understanding-ai-more-than-just-1s-and-0s/>.

¹⁴ *About Access Copyright*, ACCESS COPYRIGHT, [xhttps://www.accesscopyright.ca/about-us/](https://www.accesscopyright.ca/about-us/) (last visited Dec. 17, 2024) (defining itself as a non-profit "collective voice of creators and publishers in Canada").

¹⁵ *See generally infra* note 18.

innovation. It's a market leader in offering licensing solutions on a collective basis. So, the terms "blanket" and "collective" are used interchangeably. CCC was created to deal with a really disruptive technology at the time that today, of course, we don't think of as disruptive anymore, and that was the photocopier.¹⁶ At the time, there were a lot of concerns about the harm that photocopiers might cause to the publishing industry.¹⁷ So the CCC¹⁸ initially created direct licenses, one-to-one licenses, but it quickly evolved that model to collective licenses because that was a much more efficient way of streamlining the rights and licensing them out, to the benefit of the publishers, the rights holders, and, of course, the users as well.¹⁹ The licenses that CCC offers are all voluntary, non-exclusive licenses.²⁰ They are opt-in for the rightsholders.²¹ So, it started off with photocopy (paper-to-paper copying), then it evolved to also include digital (paper-to-digital and digital-to-digital copying), and as of July of 2024 it also includes AI use rights.²²

Phil Sherrell: And you mentioned to me when we were talking earlier, Roanie, it's not a new license, the GenAI license, it's more of an addendum to the existing license?

Roanie Levy: With the AI license, we don't actually limit it to generative AI: is an extension of the core annual copyright license that CCC offers. And this addendum has been added to all corporate licenses.²³ We have over 2000 corporate licensees that have been relying on the annual copyright license, or the ACL, in order to share and store content for over 40 years, like I said, initially with

¹⁶ The Xerox 914 - the first automatic, plain-paper commercial copier - was announced in 1959. *Xerox History Timeline*, XEROX, <https://www.xerox.com/en-us/about/history-timeline> (last visited Dec. 17, 2024).

¹⁷ This was especially the case with STM (scientific, technical, and medical) journal content. *See, e.g.*, *Texaco v. American Geophysical Union*, 516 U.S. 1005 (1995).

¹⁸ The Copyright Clearance Center was formed as a not-for-profit in the wake of the Copyright Act of 1976, which did not contain any provisions regarding large-scale photocopying, after a recommendation from Congress that publishers "work out means by which permissions for uses beyond fair use can be obtained easily, quickly, and at reasonable fees." H. Rept. 83, 90th Cong., 1st Sess. 33 (1967).

¹⁹ CCC's collective license is currently known as the Annual Copyright License. It is typically taken by corporations to cover their internal use of licensed content. For more information, see *Products: Annual Copyright License*, COPYRIGHT CLEARANCE CENTER, <https://www.copyright.com/solutions-annual-copyright-license/> (last visited Dec. 17, 2024).

²⁰ Rightsholders are free to license their content to others in whatever ways they choose.

²¹ Rightsholders must declare the content that they wish to include in the licenses.

²² *CCC Pioneers Collective Licensing Solution for Content Usage in Internal AI Systems*, COPYRIGHT CLEARANCE CENTER: PRESS RELEASES (July 16, 2024), <https://www.copyright.com/media-press-releases/ccp-pioneers-collective-licensing-solution-for-content-usage-in-internal-ai-systems/> [hereinafter "CCC PRESS RELEASE"].

²³ "The inclusion of AI re-use rights makes the [Annual Copyright License] the first-ever collective licensing solution for the internal use of copyrighted materials in AI systems." *Id.*

photocopying and then with digital.²⁴ And the license has now been extended to allow AI uses as well.²⁵

There are some important parameters around the use of content under the license for AI, and these are parameters that were and always are in the core license itself. One important parameter is that the content needs to have been legally acquired by the licensee. The second important parameter is that the use cannot substantially replace the purchasing or subscription to the content itself. And the third parameter, which is particularly important in this AI context, is that the use is for the internal business purposes of the licensee and not for public-facing uses.

Phil Sherrell: So, just to contextualize that, could you give us an example of something that a licensee can now do with the license under the AI addendum?

Roanie Levy: Yeah, that's a good question. And examples are always helpful. So, first, more generically, what can they do with the content that's covered by the AI addendum? They could use the content to train or fine-tune an LLM.²⁶ They could use the content in a RAG system, they could use the content in a prompt.²⁷ So, these are all things that people do with AI increasingly today. And the types of uses that they can do, they could use these systems in order to do market analysis. They could use it to summarize or synthesize an individual scientific journal or a database of journals. They could use the AI rights in the license to do

²⁴ CCC began offering its licenses with its launch on January 1, 1978, the effective date of the Copyright Act of 1976. *See, e.g.*, R. Bruce Rich, *History of CCC and Relevance to Texaco Case*, in *CREATING SOLUTIONS TOGETHER: LESSONS TO INFORM THE FUTURE OF COLLECTIVE LICENSING* (Copyright Clearance Center, 2021).

²⁵ CCC PRESS RELEASE, *supra* note 22.

²⁶ A large language model ("LLM") is a specialized type of artificial neural network ("ANN"), a structure that has been used in machine learning systems since the late 1950s. An ANN is a structure, modeled on neurons and synapses in the human brain, that maps inputs to outputs through a series of mathematical transformations. ANNs "learn" by receiving training data that consists of inputs and their associated outputs (e.g., images and text descriptions of them, or descriptions of houses and their actual sales prices), and adjusting the internal mathematical formulas so that new inputs result in appropriate outputs. LLMs take large bodies of text material as training data, break them down into "tokens," and ingest the tokens into the ANN so that text inputs ("prompts") result in useful outputs. *See Overview of AI and ML*, in DONNIE WENDT, *THE CYBERSECURITY TRINITY: ARTIFICIAL INTELLIGENCE, AUTOMATION, AND ACTIVE CYBER DEFENSE* (APress, 2024).

²⁷ Retrieval-Augmented Generation ("RAG") is a technique that extends the capabilities of LLM-based generative AI systems by consulting sources (e.g., websites or databases) beyond the data used to train the LLM, and typically citing those additional sources in outputs. RAG allows generative AI tools to produce output that is potentially more up-to-date than traditional LLM-based AI tools and to cite the sources consulted so that users can ascertain them or follow up with further research. *See Kim Martineau, What is retrieval-augmented generation?*, IBM (Aug. 22, 2023), <https://research.ibm.com/blog/retrieval-augmented-generation-RAG>.

information extraction, trends analysis, to accelerate product discovery or R&D within their businesses. They could use the license to do sentiment analysis.²⁸

An important concept to understand is that the license is not a use-based license.²⁹ It doesn't actually describe the uses I've just mentioned. It sets out that you can make a reproduction of licensed content in an AI system. The definition of AI is pretty broad. It doesn't limit it to generative AI. And then the license sets out what you cannot do. So, you can use the content in an AI system, but first of all, you have to legally acquire the content, the use cannot replace the purchasing or subscription, and importantly, you cannot do an external use of either the output or the LLM that you trained or any other use of the content under license.

You cannot make any of the permitted uses under the license available externally to the company. So, a pharmaceutical company, for example, can fine-tune an LLM in order to accelerate their product discovery, but then they cannot make that fine-tuned LLM available to their customers in order to query the LLM for side effects. That's just one example, but you could imagine many others.

Phil Sherrell: Thanks, Roanie. That's really helpful. Dave, you're our second panelist that's responsible for a new GenAI license, so please introduce yourself and Calliope.

Dave Davis: Yeah, so my name is Dave Davis. I'm the CEO and co-founder of Calliope Networks.³⁰ Calliope Networks was founded at the beginning of 2024. We are aggregating copyrighted audiovisual works, movies, TV series, news, et cetera, to license to AI companies for model training. My background is I spent about 15 years at movie studios doing a lot of content licensing. And then I was at the Motion Picture Licensing Corporation for three years as Chief Commercial Officer.³¹

So, we are in business today. We are aggregating content. We've aggregated a catalog of about 30,000 hours of audiovisual content today, and we're in active conversations with AI companies.

Phil Sherrell: What sort of terms are on offer, Dave? What, broadly speaking, is the Calliope license going to cover?

Dave Davis: The Calliope license will enable companies that are building video models, training video models, it allows them to use our audiovisual content for that training and fine-tuning those video models. So, for example, these models need literally millions of hours of content [for training]. The most famous

²⁸ Sentiment analysis analyzes large volumes of text to determine or summarize opinions expressed in them. For example, sentiment analysis can be used on social media posts about a book to summarize public opinions about the book. For more information on the multifaceted reality of sentiment analysis within social media platforms, see Margarita Rodríguez-Ibáñez et al., *A review on sentiment analysis from social media platforms*, 223 EXPERT SYSTEMS WITH APPLICATIONS 119862 (Aug. 1, 2023).

²⁹ That is, it does not specifically enumerate uses that are permitted under the license or any consideration for such uses.

³⁰ *Dave Davis, supra* note 1.

³¹ *Id.*

of these video models are text-to-video models like Sora³² and Runway,³³ and people can enter into a prompt window anything they want to generate a video of. If you want to generate a video of a cowboy riding a camel through the streets of Tokyo, for that model to generate those things, it needs to have had prior exposure to cowboys and camels and Tokyo. And then if you change your mind and you don't want a camel but you want a unicorn, and you don't want Tokyo but you want New York City, it needs to know those things, too. So, what they need is nearly infinite. It takes a lot of data.

Phil Sherrell: And so, I think it's clear from what you've been saying that the permissions available under your license include external use as well.

Dave Davis: Yeah, for sure. Yes. Yeah, definitely external.³⁴

Phil Sherrell: Yeah. And are you able to talk about who the rightsholders are that are in Calliope, or is that still under wraps?

Dave Davis: Well, it's not confidential, but we don't talk too loudly about the rights holders. We're working on behalf of more than 30 film and television distribution companies and production companies. And I talk to probably five to 10 companies like that a week. The companies we're working on behalf of come from all over the world. We worked on behalf of American companies, Canadian companies, Latin American companies, European, Asian. We have content from India and from Malaysia, Sub-Saharan Africa, et cetera. And we think one of the selling points of our license to AI companies is that our catalog is globally diverse and really rich.

Phil Sherrell: And Dave, you've come with a special treat for us today. We've got some breaking news from you. I said earlier there were two GenAI licenses on this panel,³⁵ but Dave's about to launch a third. Hold your breath.

Dave Davis: That's true. Yes. Our core license, what I've just described to you, is really a direct license. It doesn't quite fit into the blanket licensing theme. These are individually negotiated licenses, and they may not cover the full breadth of the content in our catalog. We are announcing today a new license that we've provocatively titled License to Scrape.³⁶ We're going to work with YouTube

³² See generally SORA AI, <https://www.soraai.onl/> (last visited Dec. 17, 2024).

³³ See generally RUNWAY, <https://runwayml.com/> (last visited Dec. 17, 2024).

³⁴ That is, companies that take Calliope's license may use the covered content for tools and generated content that they distribute to their customers or to the public. See *CCC Launches Collective AI License*, COPYRIGHT CLEARANCE CENTER: BLOG ON ARTIFICIAL INTELLIGENCE (July 25, 2024), <https://www.copyright.com/blog/ccc-launches-collective-ai-license/>.

³⁵ Referring to CCC's AI-enhanced Annual Copyright License and Calliope's core direct licenses. See *Products: Annual Copyright License*, *supra* note 19.

³⁶ "Scraping" refers to copying content from an online service, in this case YouTube. See Kate Knibbs, *This Startup Wants YouTube Creators to Get Paid for AI Training Data*, WIRED (Sep. 30, 2024), <https://www.wired.com/story/license-to-scrape-youtube-ai-data-license-creators/>.

creators, and they can opt into our license.³⁷ And then, when large AI companies want to come somewhere for a large basket of copyrightable content that they're scraping online, and many of these AI companies are scraping content online, as most of you will know, they can come to us for what will become a basket of millions of hours of content. You heard it here first.

Phil Sherrell: It's very kind of you to save that for us. Thank you. And so, that license will have a sort of similar function to the previous one you described in enabling the training of models and then the proliferation of content?

Dave Davis: Yeah, absolutely. Now, the nice thing operationally is it's very simple. So, for our direct licenses, we're delivering MP4 files generally speaking. For this license, the AI companies can kind of do what many of them are used to doing, which is continue to scrape that content. There's no file delivery. There's pretty simple reporting, et cetera. It's very simple.

Phil Sherrell: Thank you, Dave. Yes. Let's turn to the position of the Authors Guild, and I'd like to introduce Umair Kazi. Umair.

Umair Kazi: Thanks, Phil. I was going to push the blanket analogy a step further and say blankets can also be bundled. I'm Umair Kazi. I'm the director of policy and advocacy at the Authors Guild.³⁸ For those of you who don't know the Guild, it's the largest and oldest organization of professional authors. Since 1912, the Guild has advocated for the rights of authors in areas ranging from copyright and fair publishing terms to antitrust and free expression and other areas. We have been an active participant in the AI policy discussions now for a number of years, representing authors and our peers in the creative industries. My role at the Guild encompasses copyright policy, a lot of AI work, some work on labor, antitrust, and free expression issues as well. And I previously served as the Guild staff attorney, so I looked at a lot of publishing contracts. So, if anyone wants advice on their publishing contract, come to the Guild. Do you want me to go into the Guild's position on these issues?

Phil Sherrell: Yes please.

Umair Kazi: Since a little before ChatGPT came online, but around the same time, we've emphasized a couple of points that I think you'll hear repeated over and over again. But frankly, I think that they need to be repeated. It is that no AI use should occur without the creator's consent and compensation. And sometimes, the people add another C or two C's to that list and say credit and control. But I think consent is broad enough to encompass credit, and attribution. Control essentially means determination of the terms of licensing. So, when we talk about the blanket, maybe the creator should have a say in the design of whether they want something floral or something more geometric. It's just, I can't let go of the blanket now.

Phil Sherrell: We need to keep up the blanket analogies.

³⁷ That is, they can offer their YouTube videos for inclusion in Calliope's License to Scrape, such that if an AI platform copies the videos for use in AI training. *Id.*

³⁸ *Umar Kazi, supra* note 3.

Umair Kazi: We know that the existing models were not trained on licensed text, and that's why we've been lobbying for laws to implement more transparency around what the training data was.³⁹ And we'll talk a little bit more about that in the panel. But we also recognize that AI, especially LLMs, are here to stay. We can't put the genie back in the bottle. There are a number of lawsuits ongoing right now on whether or not training is fair.⁴⁰ I think they're going to come down to the question of fair use.⁴¹ Essentially, I think everyone here can agree, but there's infringement lawsuits ongoing right now.

But even if courts find that training of existing models is infringing [and not covered by fair use], the LLMs will continue to be trained and developed, and it's unlikely that any court will order them destroyed. So, in our view, that leaves the path forward, which is the path that we believe that AI companies should have taken from the beginning, is to license works from authors and appropriate rightsholders. I certainly see it as a positive sign, and in many ways, for the last year and a half, I've been alongside a lot of other creator organizations in DC and having these conversations, I think it's a testament to the advocacy of creating and creator groups that we're already seeing a lot of movement in that direction.

I mean, we just heard examples of two novel licensing arrangements for AI uses. And in the last few months, the Guild has been in conversations with companies who are building similar platforms and licenses for books and other text content. We give them feedback on how to make the licensing terms, features, and their product design fair and equitable for authors. And our primary interest, again, is to give authors control of the licensing and the terms of licensing, which also includes whether or not the author wishes to allow certain types of output uses. We've discussed technologies that can mask specific attributes of a text in the training process so that the output cannot be prompted to, say, produce a summary or competing work, or a derivative work. So, those are really the kinds of uses we are extremely concerned about that dilute the market. And, we're encouraging the development of licensing platforms with the goal to let many flowers bloom.

³⁹ For example, Generative AI Copyright Disclosure Act, H.R. 7913, 118th Cong. (2024).

⁴⁰ These include, for example: *Alter v. OpenAI, Inc.*, No. 23-CV-10211 (S.D.N.Y. Filed Nov. 21, 2023); *Ctr. for Investigative Reporting, Inc. v. OpenAI, Inc.*, No. 24-CV-4872 (S.D.N.Y. Filed Oct. 15, 2024); *Daily News L.P. v. Microsoft*, No. 24-CV-3285 (S.D.N.Y. Filed July 3, 2024); *Huckabee v. Bloomberg L.P.*, No. 23-CV-09152 (S.D.N.Y. Filed Oct. 17, 2023); *The Intercept Media v. OpenAI, Inc.*, No. 24-CV-1515 (S.D.N.Y. Filed July 15, 2024); *Kadrey v. Meta Platforms, Inc.*, No. 23-CV_03417, 2024 WL 235199 (N.D. Cal. Filed Jan. 22, 2024); *Leovy v. Google LLC*, No. 23-CV-03440 (N.D. Cal. Filed July 11, 2023); *Nazemian v. NVIDIA Corp.*, No. 24-CV-01454 (N.D. Cal. Filed Mar. 8, 2024); *New York Times v. Microsoft Corp.*, No. 23-CV-11195 (S.D.N.Y. Filed Dec. 27, 2023); *Thomson Reuters Enter. Centre GmbH v. ROSS Intelligence, Inc.*, 20-CV-00613 (D. Del. Filed Dec. 13, 2024); and all of the OpenAI ChatGPT Litigation, which can be found at Case Tracker, *supra* note 8.

⁴¹ See, e.g., *FAQs on the Authors Guild's Positions and Advocacy Around Generative AI*, THE AUTHORS GUILD, <https://authorsguild.org/advocacy/artificial-intelligence/faq/> (last visited Dec. 17, 2024).

Phil Sherrell: Thanks, Umair. So, let's move on to discussing some of the potential issues and complications that come up in this area. Let's first of all focus on the topic of coverage, then the scope of these licenses. And it's time to introduce Rachel Fertig from DLA Piper. Rachel, tell us about yourself, please.

Rachel Fertig: Thanks so much, Phil, and it's really a pleasure to be with all of you today. My name is Rachel Fertig. I'm a partner at DLA Piper and really am just focused on copyright day in and day out.⁴² My career has been, first as copyright counsel for the Association of American Publishers during Congress's initial review of the US Copyright Act,⁴³ and then I moved to the Copyright Office as a Ringer Fellow.⁴⁴ And I believe those applications are open right now and highly recommend that career path. And then I went into private practice.⁴⁵

Now, I've been at DLA Piper for a few years, really focusing on AI from across different perspectives, from the AI model developers to a number of companies looking to fine-tune or ground their models, as well as working with amazing startups in this space, trying to solve the value gap between the creators and the AI developers, including Created by Humans and their GC,⁴⁶ who is here today. There's just a lot of different interests in this market about how to actually address where we are, because I think, as many of the panelists have mentioned, there's litigation going on, but we have no answers.

And so, this is the right time for people to be looking into how to aggregate content to meet the needs of AI developers, but also the needs of creators. So, I'm happy to go into a couple of those different issues that we've been looking at. And I think, Phil, we had discussed really the two ends of the spectrum that we have on this panel, which is great, which is with CCC, an internal use-only license, which has tremendous value for enterprises to be able to use the legally acquired content that they have and mitigate their risk and exposure to potential copyright infringement. And then on the other side with Calliope Networks, the external use licensing model, which is critical to a number of companies that actually want to commercialize the outputs either for themselves or for clients of those businesses. And so, you really need both types of licenses in the market, and there's no one way to license your content for AI. In working with different rights holders, these are the types of issues that come up when people say, "Okay, I want to license my content. How do I do that?" You have many different options, and candidly, nobody has the perfect solution.

Right now, that's what makes this exciting for a copyright nerd like myself: everybody's trying to figure it out. It's a very exciting time to be trying to find

⁴² *Rachel Fertig, supra* note 2.

⁴³ *See Copyright Review Hearings, 2013-2015*, U.S. COPYRIGHT OFFICE: CONGRESSIONAL HEARINGS AND STATEMENTS TO CONGRESS, <https://www.copyright.gov/laws/hearings/> (last visited Dec. 17, 2024).

⁴⁴ *See Ringer Fellowship*, U.S. COPYRIGHT OFFICE, <https://www.copyright.gov/about/special-programs/ringer.html> (last visited Dec. 17, 2024).

⁴⁵ *Rachel Fertig, supra* note 2.

⁴⁶ *Id.*

different ways to create these new markets and solve a lot of different risks that are unknown to be able to support creators and technological developers. So, I think you see on the internal and external licensing distinction that really comes into play for pricing, and I think maybe we'll get into that a little bit more. But for external uses, often that increases the price that a rights holder is going to want for their use, because it also relates to a higher potential risk for what happens to that output if it is made available externally.

Here are a few things to think about: external versus internal uses of copyrighted works; How do I enforce [the license] if the external use doesn't stay within the license parameters that I've set for the price that I'm willing to license it? Whereas with internal uses, you have a potentially easier mechanism for enforcing the limitations of those licenses. The price will depend on the rights conveyed and the enforceability of the terms. And so, you may be able to lower the price, but get more people to take the license because it does have an easier entry point for actually licensing that content. So, we can go into all the myriad of different guardrails and risk factors that go into putting together these licenses. I'd certainly open it up to other panelists too to talk about their experience with that.

Roanie Levy: If I could just jump in. I think it's also important to highlight that there isn't and likely will never be a one-size-fits-all [license]. I think there is going to be a multitude of licenses and a vibrant rights market to address the different situations depending on who is actually using the work and for what purposes. For instance,, is it a license for internal business purposes? Is it an LLM operator for training purposes? Is it an AI provider in order to create products that are then public facing? And with every given context, there's also going to be a multitude of different solutions from direct licensing situations to collective licensing So, the one-size-fits-all, we're going to have one license that solves all the problems, I don't think it's on the horizon.

Phil Sherrell: Dave, just going back to that point about external [licenses], how did you persuade rights holders to come on board with the idea of liberating these LLMs to go off and potentially compete with their content in the future?

Dave Davis: As part of Calliope Networks, I talk to a lot of rights holders every week and they run the gamut. The entertainment industry is under threat generally and has been in decline for the last 10 or 15 years from all kinds of threats. These tend to be businesses that are trying to find incremental revenue. Their revenues are declining a few percent per year. And they run the gamut from people who are very close to the art, to people who are sort of just very economically pragmatic. For those who are closer to the art, it's sometimes a more direct and difficult conversation.

People who are pretty economically pragmatic I'll tell them, "Listen, this is an incremental revenue stream. There's no actual exposure of the content in the traditional sense. It's not going to go up on a television screen or on a movie screen or something, and you get paid money for it." And for some, that's enough. For those who are a little more concerned about the risks of AI, it's a deeper conversation. But one of the big things I'll tell them is, "You've basically got

three choices. 1) You can ignore AI,” which many of them candidly would like to do. 2) “You can litigate,” which maybe for some, that’s the right course, or, 3) “You can engage.”

And I’m advocating engagement. I spent a lot of my career at Paramount.⁴⁷ Paramount was part of Viacom.⁴⁸ We spent a lot of time choosing a litigation path with YouTube back in the day, which Dan will remember in the back.⁴⁹ And at the end of all of that litigation, candidly, Viacom had kind of lost relevance relative to the MTV generation.⁵⁰ And so, through my lived experience, I think engagement really has a place in this market. And that, again, some distributors, some producers aren’t ready, and that’s okay. And some are. And so, we’re engaging with those who are.

Rachel Fertig: I think, just picking up on that--with engaging with people where they are-- we’re seeing in the licensing space that because there is a spectrum of fear and excitement about the opportunities and risks of AI, there are different types of licenses that some rights holders may be more comfortable with than others. So, I think, Roanie, you mentioned either training or fine-tuning an LLM versus use with grounding with RAG.⁵¹ And so, certain licenses that we’ve worked on will have a training authorization, and that may come with a higher price point and higher risks because there’s not really standard methods of unlearning from LLM models.⁵²

Whereas rights holders that really just want to dip their toe in the water and see if it’s going to be worth their while may want to engage in a license that only allows for RAG or grounding uses⁵³ and maybe for a limited amount of time. And so, then if they don’t actually like having their content involved in certain types of outputs from AI models, they can choose to terminate the license or not renew it, and not have their works forever integrated into a training data set for a particular model. So, there are a number of different ways for rights holders to engage in the licensing space. And to your point, Roanie, there’s not a one-size-fits-all approach.

⁴⁷ *Id.*

⁴⁸ Currently known as Paramount Global. *See About, PARAMOUNT*, <https://www.paramount.com/> (last visited Dec. 17, 2024).

⁴⁹ Daniel Cooper, current President of the Copyright Society and Senior Vice President and Deputy Chief Counsel, Intellectual Property, NBCUniversal; former SVP Intellectual Property, Paramount Pictures. *See Daniel Cooper, THE COPYRIGHT SOCIETY*, <https://copyrightsociety.org/bio/daniel-cooper/> (last visited Dec. 17, 2024); *Viacom Int’l, Inc. v. YouTube, Inc.*, 676 F.3d 19 (2d Cir. 2012).

⁵⁰ YouTube took over from cable television as a popular destination for video content.

⁵¹ Martineau, *supra* note 27.

⁵² That is, removing certain training data from an LLM so that the LLM acts as if it had never been trained on that data.

⁵³ This is in reference to a license that only allows use with retrieval-augmented generation so that sources will be cited in the output, as opposed to standard generative AI use which typically does not identify sources in its output. Martineau, *supra* note 27.

And right now, there are just many different experiments with different types of AI licenses and the different types of outputs that you may be willing to authorize and what price points go along with them. And each one of those comes with a whole spectrum of different guardrails and pricing considerations. So, full lawyer employment act, I guess, at this point, but it's a very dynamic and interesting space to be in.

Roanie Levy: The only thing I would add is to highlight that the emotions around AI are changing really, really rapidly. We saw a big change from when we first started engaging with our publishers and with our corporate customers in January 2023. So, just a couple of months after the public release of ChatGPT compared to the conversations today and the comfort level today. And it's not that long of a time, and it's changing really, really rapidly. Part of it is just that there's a bit of denial at first, and then acceptance. So, there's some acceptance that this is not going away and we need to engage. There's a comfort level that is increasing as well. So, things are changing really, really rapidly, even from just the emotional side of how to address these issues.

Phil Sherrell: Umair, are the authors feeling comfortable with their content being licensed for external uses?

Umair Kazi: I think there is still a lot of trepidation, but I think there's also intrigue, especially from independently published authors, who form a large segment of the author community now, who tend to be more entrepreneurial. who are kind of are curious as to what sort of potential derivative uses might come, and whether this will open up income streams.

But the experience that authors have had with new technology or new sort of models of distribution and commercializing work has been quite negative. So, what ends up happening is that a new kind of mode of distribution or use may come around, and there may be a promise to it, but the author may sign away rights over broadly and then cut themselves when the market finally starts maturing, and then the authors can clearly see the opportunities that they missed. From the beginning, there needs to be enough granularity to the licenses. And I was thinking about the idea that is some sort of blanket [license]. Maybe it's something uniform, or perhaps a bundle where there can be variances within the licensing terms. I think that the main concern that authors have is that without any kind of guardrails, competing works will just flood the market.⁵⁴ There's also other concerns. For instance, the way the publishing economy works now, especially with ebooks, is that if you're getting paid out of a subscription pool, how much does the sort of glut of AI-generated works lower the value of your work in that? There are all of these economic questions that we're thinking about.

Phil Sherrell: We'll come back to the money in a minute. But before we do, I just want to touch on retrospective use. And I guess it'd be interesting to hear from Roanie and Dave whether in their licences they are licensing the sins of the past that the LLMs have committed in ingesting works without permission.

⁵⁴ For example, AI-generated works in the same style as that of a given author or series, or works that use characters from existing works.

Roanie Levy: So, for practical purposes, there is a waiver in the license, but the waiver for the sins of the past, as you put it, is to the benefit of the licensee only. So, the corporate customer that is using an LLM that is potentially or might be found infringing a given work, the rights holder is waiving any rights, vis-à-vis that corporate licensee of CCC. So, there is that waiver. The other waiver as well that's in the license is with respect to TPM (Technical Protection Measures). So, there is a waiver for the removal of TPMs, whether it's for the sins of the past of the LLM operator or the actual corporate licensee who is removing TPMs in order to use the content in an AI system.

Phil Sherrell: And Dave, is Calliope in the business of retrospective licensing?

Dave Davis: Yeah, especially for our license to scrape. We will retroactively forgive past acts. We don't want to create a barrier to negotiation with people fearing that negotiating and licensing for the future will be an admission that they did something wrong in the past. We'd just like to get to the future state where people are licensing.

Phil Sherrell: And then I guess the question that follows, and this will take us onto the money, is there some sort of lump sum for payments for the past, I guess, particularly in the case of a license direct to the GenAI companies?

Dave Davis: For our direct license right now, we're very public about the model and even about our pricing, in part because this business is so new and the handful of deals that were being done out in the marketplace were so chaotic and so all over the place. And there really weren't very many deals being done. There was a mismatch of expectations with content owners thinking their content is incredibly valuable and AI companies thinking it's not very valuable at all. We needed to bring those things together. We've been very public about that for our direct license. We have a list price at \$6.25 per minute for HD content.⁵⁵ And we can talk about the parameters of how you derive a price in a market where there is no price.

Phil Sherrell: And that's included within the \$6.25 per minute, that includes the past use as well?

Dave Davis: That would include past uses, but for the direct license, in most cases, the AI companies haven't found that content. That's much more important for YouTube-based content.⁵⁶ But if it's a film from Malaysia, it's probably not already in their data set.

Phil Sherrell: And I guess, Rachel, in terms of what you're seeing in the market, I'm guessing that to get a license signed in this space, it's going to have retrospective use covered from a developer perspective.

Rachel Fertig: I guess I'm going to be a little contrarian on that. I don't think necessarily it has to have the retrospective waiver and release in it because there

⁵⁵ High Definition, which typically refers to video with a resolution of at least 720x1280 pixels.

⁵⁶ Because YouTube content has presumably been scraped, whereas content that is only available offline has not.

are a number of, I would say, enterprise users who are looking to customize existing models, and so they may not have the same concerns about past exposure. And it's really on a go-forward basis. They're looking to take a license as well as additional startups getting into the market and starting with clean training sets. . In those cases, you don't have the past use. So, I think it really just depends on what type of AI development you're doing if the retrospective release is necessary. I also think in some circumstances you may not be able to get it from the rights holders that you're getting licenses from.

In the US, in a lot of co-ownership situations, you may be able to work with a single rights holder who has rights to grant a non-exclusive license, and we're seeing most of the AI licenses are non-exclusive grants. And so, they can only release their own claims unless they have a separate written agreement with their co-owners that they could release their accrued claims as well. So, you may not practically be able to get a full release for certain of these non-exclusive AI licenses in the first place.

Phil Sherrell: And then just in terms of the flow of money then, Dave, you've talked about the amount that's being paid per minute. Are you able to tell us a little bit about where it goes once it's been paid to you?

Dave Davis: At Calliope, we work on a revenue share basis with our partners. The bulk of the money, the vast majority, goes back to our partners, and it's simple to allocate because if they gave us 10 hours of content or 100 hours of content, we just straight-line that pricing relative to them.

Phil Sherrell: And Umair are you expecting ultimately the authors to then see money trickle back down through the royalty flows?

Umair Kazi: Yeah, absolutely. I think pricing is the million or whatever dollar question. We don't even know how big the market is. But it is that question. What I am interested in is pricing that also allocates for the use. We talked about it. I think these notions, these concepts are getting a little bit more reified as we have the conversations. But when we talk about these licenses, we're kind of talking about a universe of uses, a lot of uses that we haven't even discovered yet.

So, oftentimes I'll get a report from an author. I'll just give one example. We have members that wrote up an emotional thesaurus, so for writers to sort of like figure out what characters are feeling.⁵⁷ And I've seen now two chatbots built on that, and the chatbots charge \$3 a month.⁵⁸ So, they basically chat-bottified this book, and someone's earning \$3 a month from subscriptions. So, there are these uses that we haven't contemplated yet. And the licenses will have to sort of have some sort of future-proofing in terms of that.

But the other question that I wanted to say is that training is something a little bit more discrete. You can say, "Okay, for a lump sum you can train on my book." That's it. But what happens after that? Training is not also entirely distinctive from the output. You can put some restrictions on the output, but the model is still

⁵⁷ ANGELA ACKERMAN & BECCA PUGLISI, *THE EMOTION THESAURUS: A WRITER'S GUIDE TO CHARACTER EXPRESSION* (2nd ed., 2019).

⁵⁸ These chatbots are no longer available.

kind of pulling on that information. And so, one of the things that we're also talking to some providers about and exploring is the idea of when a chatbot produces an output, how a token value for the work that is in the bundle license is calculated and then payment being readjusted. For instance, someone like George R. R. Martin's work, just because it's so popular, it's going to be used more likely in the chatbot content.

Phil Sherrell: It sounds like we're going back to the world of blockchain and smart contracts, which was the buzzword five years ago.⁵⁹

Umair Kazi: Yes. Absolutely. Yeah. But so, there has to be some flexibility.

Dave Davis: I was just going to add, there are a lot of companies working on pro rata attribution models. I think some are calling it computational attribution, these attribution models.⁶⁰ There's a company in LA called ProRata.AI that was just founded to do that.⁶¹ There's an AI company out of Israel called Bria.AI that's doing that.⁶² And I think NVIDIA with Getty is doing that as well.⁶³ So, there are various experiments on that. It seems really hard to me. What we're doing is a lot simpler.

Phil Sherrell: And just to wrap up on pricing, Rachel, do you think there's a bit of a fear in the market around the risks of setting the bar somewhere which might then prove to be too high or too low, but it does set a bar for litigation, damages purposes, and so on?

Rachel Fertig: I think there is that for certain rightsholders, for sure. But I also think we're all talking about voluntary licenses, and so there's going to be negotiation. So, to the extent that licenses are entered into, they're going to be at whatever it is mutually agreeable at this point, given where we are with uncertainty under US law. And so, I think to your point earlier, Dave, a lot of rightsholders, maybe a year ago, were trying to ignore AI. But now you're seeing that this is here to stay. And so, if you're not getting in the market now, I think there's a lot of rights holders coming around to thinking we have to just actually jump in and start experimenting, even if it's not perfect. But yes, you're always worried about what you're leaving on the table.

Phil Sherrell: Yeah. And you were saying earlier that when a collective license is formulated and offered in the US, there's no ability for the licensee to

⁵⁹ See, e.g., Rosenblatt, Bill, *The Future of Blockchain Technology in the Music Industry*. 66 J. COPYRIGHT SOC'Y 271 (2019).

⁶⁰ In this case, the term refers to models for determining the text that was used in training a large language model that led to a specific output and the extent of its use.

⁶¹ *About Us*, PRORATA.AI, <https://www.prorata.ai/about> (last visited Dec. 17, 2024).

⁶² *About Us*, BRIA, <https://bria.ai/about-bria> (last visited Dec. 17, 2024).

⁶³ *Overview and Pricing*, GETTYIMAGES: AI GENERATOR, <https://www.gettyimages.com/ai> (last visited Dec. 17, 2024); *NVIDIA Brings Generative AI to World's Enterprises With Cloud Services for Creating Large Language and Visual Models*, NVIDIA: PRESS RELEASE (Mar. 21, 2023), <https://nvidianews.nvidia.com/news/nvidia-brings-generative-ai-to-worlds-enterprises-with-cloud-services-for-creating-large-language-and-visual-models>.

go to court and challenge the price unless it's one of those special music licenses that has the CRB.⁶⁴

Rachel Fertig: At least everything that we've been discussing, these are all either startups or other organizations that aggregate content privately through voluntary agreements. I think the general preference in the US is to do things between privately negotiated agreements as opposed to having something imposed, particularly where we're all still figuring out the pricing. So, yeah, there really isn't a mechanism for the majority of this to go to court and challenge those prices.

Phil Sherrell: It may be a little different in the UK and the EU, where most countries have freestanding rate courts which can rule on the pricing in any sort of collective license.⁶⁵ So, I think there may be some push to challenge rates, although only come if the rightsholders are setting the bar too high, I guess, at the beginning.

Umair Kazi: The reference in the US has been for voluntary licensing. I think there's a suspicion of collectives in the US.

Phil Sherrell: Rachel, in terms of catalysts for change, we don't want to get too bogged down in the detail of fair use, which is obviously a much litigated topic in this space. But we talked about the impact that the existence of blanket licenses might actually have on those disputes and the ability of the developers to rely on fair use defenses.

Rachel Fertig: Yeah, and just to reiterate Phil's disclaimer at the beginning of this, these are my personal views, and not on behalf of DLA or any of our clients. As many of you know, there's so much fair use case law in the US. You can essentially find something for either position that you're on. So, I think just sort of at a high level, I think on the AI developer side, there's plenty of case law to cite to say that even with the development of licensing markets, uses that would otherwise be found under the first factor of the fair use test to be transformative don't necessarily change the outcome of the fair use calculus in those cases.⁶⁶ On

⁶⁴ The Copyright Royalty Board, established pursuant to the Copyright Royalty and Distribution Reform Act of 2004, is a panel of three Judges within the Library of Congress that sets royalty rates for statutory licenses such as those provided for in 17 U.S.C. §§ 114 (for sound recording performances via digital radio) and 115 (for reproductions of musical works). See generally *About Us*, U.S. COPYRIGHT ROYALTY BOARD, <https://www.crb.gov/> (last visited Dec. 17, 2024).

⁶⁵ See, e.g., Copyright Tribunal, UK GOVERNMENT: ORGANIZATIONS, <https://www.gov.uk/government/organisations/copyright-tribunal> (last visited Dec. 17, 2024); See also The Arbitration Board under the Act on Collective Management Organisations, GERMAN PATENT AND TRADE MARK OFFICE, https://www.dpma.de/english/our_office/about_us/further_duties/cmso_copyright/arbitration_board_under_the_cmo_act/index.html (last visited Dec. 17, 2024).

⁶⁶ For example: *Bill Graham Archives v. Dorling Kindersley Ltd.*, 448 F.3d 605, 615 (2d Cir. 2006) (“[A] copyright holder cannot prevent others from entering fair use markets merely ‘by developing or licensing a market for parody, news reporting, educational or

the other hand, if you are sitting in the content owner's side of the equation, you look to cases like *Texaco* from the Second Circuit, which generally said that you should be more fair if there is not a ready market to pay for it and less fair if there is a ready market to pay for it.⁶⁷ So, I think regardless of where you sit on the outcome of whether training should be fair use or not, I think the development of a licensing market is going to impact all of the litigations because it will become something that everybody has to grapple with if you are asserting a fair use defense once there actually is a ready mechanism to pay for the type of use at scale. We'll see how this plays out. But I think there's certainly precedent on both sides that people are going to be able to rely upon to say that it should or shouldn't change the outcome of these cases.

Phil Sherrell: That's interesting, I don't know if the rest of the panel think that's a factor in the mind of rights holders as these licenses get formulated. Is it almost pushing them to accelerate the granting of licenses in order to improve the position of rightsholders in the litigation?

Roanie Levy: I think it's one of many factors that rights holders are considering. Whether it's what happens in the courts with a fair use analysis, or what happens in the minds of policymakers and government legislators, in terms of looking at whether there needs to be exceptions or other limitations introduced to allow innovations to take place. So, that is all part of what's going on in the minds of rightsholders. And also just generally, that even with a court decision, it will probably not answer all the questions that are out there for all of the use cases. So, it's unlikely, even with 30 cases, it's unlikely to answer all that we need answers for. And there is a patchwork situation. These [AI] products are global. They're used everywhere, and the laws are different in every country. And so, finding a solution that could cross borders, that can harmonize to some extent the terms of use around copyright-protected work is really the way to go.

Rachel Fertig: I think it's also going to matter what the particular use is, to your point, Roanie. It's not just whether training an LLM or training a diffusion model is fair use, but it's also what types of outputs do you have from a particular model that are going to come into this interplay between the first factor, the purpose of the use, and the fourth factor, whether there is a cognizable market harm with respect to that particular purpose in use. So, this will continue to be a case-by-case issue.

Umair Kazi: And I'll just say that we are talking about the evolution, I think, in some ways, where we've been seeing development of the market, but I would

other transformative uses of its own creative work. [C]opyright owners may not preempt exploitation of transformative markets.”) (quoting *Castle Rock Ent., Inc. v. Carol Pub. Grp., Inc.*, 150 F.3d 132, 146 n.11 (2d Cir. 1998)); see also *Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 99 (2d Cir. 2014) (“[U]nder Factor Four, any economic ‘harm’ caused by transformative uses does not count because such uses, by definition, do not serve as substitutes for the original work.”).

⁶⁷ *Am. Geophysical Union v. Texaco, Inc.*, 60 F.3d 913 (1994).

say we're talking about the evolution of the licensing market because as CCC they've had these sorts of licenses for decades now.

Dave Davis: Everybody should go to Calliope Network's website.

Umair Kazi: It's Dave's TED Talk. Yeah. And I was just going to say that I think that the acceleration in licensing is also driven by the kind of uses that we're seeing. So, we're already seeing books coming out that are AI-generated, and the uses are, like I said, chat-botification of a book.

Phil Sherrell: And I guess the other catalyst which we spoke about for change when we were preparing for this session was the transparency agenda. You might be aware that in the EU AI Act, AI model developers are going to be required from next year to identify the works that were used to train their models. And I think that's going to have quite a catalyzing effect, because from a perspective of a copyright litigator, if you are obliged to give a rights holder a long list of works that have been used, that's quite an advantage when it comes to considering litigation or in terms of bargaining power in a negotiation, because quite often it's very hard to know exactly whether a work's been used to develop an LLM. And I think there's a transparency agenda afoot in the US as well - Rachel, you were telling us there's some breaking news from Friday.

Rachel Fertig: So, over the weekend in California, Governor Newsom signed AB 2013, which is going to require disclosure of a summary of what information was used in training data sets for models that were available for public use to Californians going back to January 1, 2022, including if any of that content has copyrighted content in it and whether or not it was licensed and when it was gathered.⁶⁸ The transparency agenda, at least at the state level, is moving on that. The US Congress has also introduced legislation to do something similar, but that hasn't moved forward, and Congress has gone on recess at this point.⁶⁹ But it's certainly something that the US is developing as well.

Phil Sherrell: And does that have a similar provision to the EU whereby if you're offering a model into California from wherever in the world, you have to comply with California law?

Rachel Fertig: From my reading of it, you do. As long as it's made available to Californians, then it would apply. But we'll see early days on all of those things.

Phil Sherrell: Yeah.

Dave Davis: I think that's kind of an important point. In the media world, we spend a lot of time licensing movies and TV series by territory. You really only have to worry about the laws of France for this license and the laws of the UK for another license. There is no concept like that for AI. Effectively these AI-models are all global. So, this is both my view and Calliope Network's view, is that it's much easier to just sign a licensing agreement than to try to harmonize some exceptions through more than 100 different copyright regimes.

⁶⁸ Generative artificial intelligence: training data transparency, A.B. 2013, 2024 Leg., Reg. Sess (Cal. 2024).

⁶⁹ AI Foundation Model Transparency Act, H.R. 6881, 118th Cong. (2023).

Questions from the Audience

Speaker 1: Thank you. You haven't talked too much about compensation for users. Mr. Davis mentioned one, and that seems strictly tied to training. So, my question is kind of multi-part or multifaceted as the problem is, which is, are the licenses you're talking about just for training? And would, for example, an author have a separate claim for infringement if an output was infringing? So, how far through do those rights go? You did allude to the problem of how to tie those, the George R. R. Martin rights. And so, for example, that's what I'm looking for, is where do these licenses go? Do they just go to training? Do they also include output? And if it is output, where's the outer limit? So, for example, I can't say, please reproduce an exact work and assume I'm going to be licensed for that. So, how do you draw the line on where the outer scope of the rights are that you're granting?

Dave Davis: So, our licenses are for training and for outputs.⁷⁰ We have certain restrictions in our license agreements that include you're not allowed to create a name, image and likeness of an actor who's in the work, you're not allowed to output an exact copy of a frame or a scene; and you're not allowed to create derivative works. Now, those are broad and subject to some interpretation, and we've left it there for now, but we are trying to address some of those concerns. From a compensation standpoint back to rights holders and participants, again, we share the lion's share of our revenue back with the company that owns the copyright.

So, the producer or distributor may have back-end deals with writers, actors, directors, but at a minimum, by enhancing their green light model, their financial model, for how much their content is worth, it will enable more content to be created and should, at the margins, create more jobs.

Umair Kazi: I can maybe just pitch in from an author-publisher point of view that there's a very strong sort of reticence for output uses because publishers and authors haven't seen what those uses could look like. So, right now, the bigger curiosity is for training, and part of the license is how much of the output use can we restrict until we have a more tangible idea of the value.

Rachel Fertig: And I would just add to that, in the licenses that I've been working on, there are lots of different specific licenses that would say training for internal use only or training for external use, and then put very specific guardrails on what those outputs from those models can include or not include. If it's a text work talking about how much verbatim content from the original material can be included in them, running them against a process to check for that content. But also with the RAG licenses, those aren't necessarily changing model weights or actually training a model.

Those are in a separate system that can be used to create outputs that are more tailored to the particular content, but also can have guardrails put in place on them

⁷⁰ That is, they cover usage of the licensed content for both training AI models and for the outputs of generative AI systems.

as well so that you do have more control. I've also heard of licensees who are asking for access to the outputs to then be able to identify what was generated based on their licenses, to then have a database to track whether those outputs are staying within the bounds of the licenses that they've granted, and then to use traditional enforcement detection services about whether those outputs are ending up in the places where they weren't supposed to.

So, say you've granted a license for personal, non-commercial use, but then something ends up on a platform for sale, then if somebody has breached the scope of your license, there's also lots of audit rights that people are putting into these licenses to be able to check in on compliance. So, there's a number of different tools that rights holders are using to try to mitigate those risks of, "Is this going to go beyond the scope of what I'm trying to license?"

Phil Sherrell: Thank you.

Roanie Levy: The CCC ACL license is only for internal uses. So, external uses are not covered by the CCC ACL. So, an external use is a separate licensable event or infringement claim event, however you want to see it. We are also talking with rights holders and with users and other stakeholders to examine what other collective voluntary solutions might be needed in the marketplace.

Speaker 2: So, my question is from the user perspective, and I know Roanie has handled this in the past. So, borrowing from the metaphor, looking under the blanket, one of the problems is there are so many ways that you can license content in. And you've talked about them, and it's wonderful, and they're so flexible. And you can even yank the license, and what do you do as a user? So, hopefully, there will be an eye towards developing as part of the infrastructure a search function so that a user can understand at any given point in time whether or not the license, be it internal or external, covers what they want to do, or perhaps Roanie will tell us if that can be covered by indemnities and warranties and licenses.

Roanie Levy: Well, specifically with respect to the CCC ACL AI addendum license, there is a search function that allows the licensee to identify the works covered by the AI Addendum. So, our licensees are able to look up at the individual work level what uses they can do of that work, including whether they could use it in an AI system. So, knowing which works are covered for which uses, obviously, has to accompany the license.

Speaker 3: Hi, Ted Sabety.⁷¹ So, I have a question about who are the rightsholders and how you can depend on that. To give you an example, Soundcloud, which is the free music distribution service that pays no royalty to musicians who upload their content to it, so get your stuff off of it, in my opinion, now has a new user agreement, buried pages into it. It says that not only are you not getting a royalty for the distribution of your sound recording, but you're

⁷¹ He is the founder and Principal Attorney at Sabety+associates PLLC. *Ted Sabety*, SABETY+ASSOCIATES PLLC, <https://sabety.net/> (last visited Dec. 17, 2024).

granting Soundcloud the right to train an AI on the material that you upload to Soundcloud.⁷²

The question then for you guys is, like for Calliope, when you say, well, the rights holders, who are we talking about? Does YouTube have one of these buried provisions in its user agreement? So, you're going to go do a deal with YouTube and now have access to every YouTube user's content because they clicked on something they never read and understood that granted YouTube the right to train, and I presume there's a right to sublicense that right to you.

The question is, from a policy standpoint, what do we do about this complete mismatch of economic power between the creators who make the content and the initial distribution platforms who are granting these rights to companies like yours when, in fact, they've gotten the rights through some clickthrough agreement? And are you willing to really live on the clickthrough agreement as the basis for providing all that content? Thank you.

Dave Davis: Yeah, we aren't operating like that. So, we really are on, I think, the side that you're advocating for. We're not doing a deal with YouTube, for example, and saying, "Give us all the data." We're doing deals with individual YouTube creators. The film and television business is a little different than music because you do end up with a thing that hundreds of people worked on that is owned by somebody, and that company has a chain of title, and they can license it. So, that's who we're licensing it from. We're not going to some murky intermediary like that. So, I think we're on the same side. Yeah.

Umair Kazi: So, you're saying there may be more than one party under the blanket? In certain industries, like in publishing, for instance, the contention will be between who owns the AI rights. Is it the publisher? Is it the author? Our position has been that the intrinsic value for training comes from the creation, the expression itself, it's not the distribution of it. But these are very much the conversations that are ongoing in the industries and I'm sure also in the music industry and certainly in the publishing industry. I just couldn't help myself making that unfortunate.

Phil Sherrell: That's a great way to end. Thank you, Umair. I think you all heard the alarm, so we're out of time.

⁷² See *SoundCloud Terms of Use*, SOUND CLOUD, <https://soundcloud.com/terms-of-use> (last visited Dec. 17, 2024).